Estimation of Speaker Age

Effects of Speech Properties and Speech Material

Sara Skoog Waller

Main supervisor: Mårten Eriksson
Co-supervisor: Billy Jansson

Faculty of Human Sciences
Thesis for Doctoral degree in Psychology
Mid Sweden University
Östersund, 2019-12-16
Akademisk avhandling som med tillstånd av Mittuniversitetet i Östersund framläggs till offentlig granskning för avläggande av filosofie doktorsexamen måndag, 16 december 2019, klockan 10.00, i Krusenstjernasalen på Högskolan i Gävle, Gävle. Seminariet kommer att hållas på svenska.

Estimation of Speaker Age
Effects of Speech Properties and Speech Material

© Sara Skoog Waller, 2019-11-04
Printed by Mid Sweden University, Sundsvall
ISSN: 1652-893X
ISBN: 978-91-88947-28-4

Faculty of Human Sciences
Mid Sweden University, 831 25 Östersund
Phone: +46 (0)10 142 80 00
Mid Sweden University Doctoral Thesis 310
Till min pappa
Acknowledgements

When I first attended the introduction course in psychology, I was amazed by the theories of human behaviour, perception and cognition, and I never expected that I would one day write a doctoral thesis in psychology. I am so happy to have had this chance and I want to express my heartfelt gratitude to everyone who have made it possible. To those who believed in me when I did not myself. It has not been a straight path. I want to express my special thanks to some persons in particular:

My supervisor, Mårten Eriksson. You have been with me on this long journey ever since the first experiment in my bachelor’s essay. Saying that you have been patient would be an understatement. Thank you so much for your support and for guiding me through all of these years!

Anders Flykt and Tanja Bänziger. Thank you for your invaluable advice and help regarding recording and speech analysis. Without your generous support, those first crucial steps would have been so much more confusing and challenging for me.

My father, Tommy Waller. Thank you for inspiring me and for your endless guidance when I was struggling in the maze of statistics, especially in the beginning of my psychology studies. You never hesitated to spend long hours answering my many questions.

Linda Langeborg, in some ways my partner in crime. Thank you for being a friend and for being someone who I can turn to with research related issues, self-doubt and random life struggles.

My fantastic colleagues in the psychology group at HiG. Thank you for encouragement and friendship, and for frequently asking the uncomfortable question - when are you finishing? Well, hopefully very soon.

All who have participated as speakers or listeners in the studies, a big thank you for lending your time and contributing to this thesis.

My family - Alex, Leonie and Noah. Along the way towards a PhD in psychology, you have constantly interrupted and distracted me, and reminded of what is truly important in life. I love you to the stars and back!
# Table of contents

Abstract .......................................................................................................................... vii

Sammanfattning ............................................................................................................. ix

List of papers ................................................................................................................... xi

Introduction .................................................................................................................... 1

Voice perception ............................................................................................................. 1

Age related changes in speech production ................................................................. 4
  Changes in speech properties ..................................................................................... 5
  Changes in speech production systems ..................................................................... 6

Perception of speaker age .............................................................................................. 8
  Perceptual cues to speaker age .................................................................................. 9
  Measuring age estimation ......................................................................................... 10
  The influence of speaker characteristics and speech material on age estimation performance .............................................................................................................. 11

Age estimation in ear witness testimony ...................................................................... 12

Aim ............................................................................................................................... 16

Summary of studies ....................................................................................................... 17

Ethical considerations .................................................................................................... 17

Study 1 .......................................................................................................................... 17
  Background and aim .................................................................................................. 17
  Procedures, design and analyses ............................................................................. 18
  Results and conclusions ......................................................................................... 19

Study 2 .......................................................................................................................... 22
  Background and aim ................................................................................................ 22
  Procedures ............................................................................................................... 23
  Design and analyses of the voice data ..................................................................... 24
  Design and analyses of the age perception data ....................................................... 25
  Results and conclusions ......................................................................................... 25

Study 3 .......................................................................................................................... 26
  Background and aim ................................................................................................ 26
Procedures.................................................................................................................. 27
Design and analyses.................................................................................................... 28
Results and conclusions ............................................................................................. 28

Discussion .................................................................................................................... 31
Discussion of main results .......................................................................................... 31
The relative importance of speech rate ................................................................. 31
Vocal and perceptual effects of age disguise ...................................................... 33
Accuracy and confidence in estimation of speaker age .............................................. 34
Methodological issues and future studies ............................................................ 36
Concluding remarks ................................................................................................ 37

References .................................................................................................................. 39
Abstract

The aim of this thesis was to investigate factors related to accuracy in estimation of speaker age and the role of certain speech properties in perception and manipulation of speaker age, as well as their interaction with the speech material that the age estimates were based on. This thesis consists of three studies. In Study 1 the aim was to investigate the role of speech rate as well as the level of accuracy in estimation of speaker age, depending on linguistic variation in the speech material (read versus spontaneous speech). In two experiments, one using read speech from 36 female and male speakers in three age groups (younger: 20-25 years, middle aged: 40-45 years and older: 60-65 years old) as stimuli, and the other using spontaneous speech from the same speakers, we investigated how changes in speech rate influenced listeners’ age estimates of young adult, middle aged and older speakers. The results revealed that listeners estimated the speakers as younger when speech rate was faster than normal and as older when speech rate was slower than normal. This speech rate effect was slightly greater in magnitude for older speakers in comparison with younger speakers, suggesting that speech rate may gain greater importance as a perceptual age cue with increased speaker age. This pattern was more pronounced in Experiment 2, in which listeners estimated age from spontaneous speech. Faster speech rate was associated with lower age estimates, but only for older and middle aged speakers. Taken together, speakers of all age groups were estimated as older when speech rate was decreased, except for the youngest speakers in Experiment 2. The absence of a linear speech rate effect in estimates of younger speakers, for spontaneous speech, implies that listeners use different age estimation strategies or cues (possibly vocabulary) depending on the age of the speaker and the spontaneity of the speech.

Study 2 investigated how speakers spontaneously manipulate two age related vocal characteristics (fundamental frequency and speech rate) in attempts to sound younger versus older than their true age, and if the manipulations correspond to actual age related changes in fundamental frequency (F0) and speech rate. The study also aimed at determining how successful vocal age disguise is by asking listeners to estimate the age of generated speech samples and to examine whether or not listeners use F0 and speech rate as cues to perceived age. Participants from three age groups (20–25, 40–45, and 60–65 years) agreed to read a short text under three voice conditions. There were 12 speakers in each age group (six women and six men). They used their natural voice in one condition, attempted to sound 20
years younger in another and 20 years older in a third condition. Sixty listeners were exposed to speech samples from the three voice conditions and estimated the speakers’ age. Each listener was exposed to all three voice conditions. The results indicated that the speakers increased F0 and speech rate when attempting to sound younger and decreased F0 and speech rate when attempting to sound older. The voice manipulations had an effect on age estimation in the sought-after direction, although the achieved mean effect was only 3 years, which is far less than the intended effect of 20 years. Moreover, listeners used speech rate, but not F0, as a cue to speaker age. It was concluded that age disguise by voice can be achieved by naïve speakers even though the perceived effect was smaller than intended.

In Study 3 the aim was to study confidence and accuracy in estimates of speaker age and whether confidence can serve as an indicator of estimation accuracy. Two experiments were performed investigating accuracy in estimation of speaker age, as well as the listeners’ confidence that their estimates were correct. In Experiment 1 listeners made age estimates based on spontaneous speech while in Experiment 2 the estimates were based on read speech. The purpose of the study was to explore differences in accuracy and confidence depending on speech material, speaker characteristics (gender and age) and listener gender. Another purpose was to examine the realism in the listeners’ confidence ratings in estimations of spontaneous versus read speech. No differences in accuracy or confidence were found due to speech material type. Although accuracy was higher in estimates of male speakers, confidence was higher in estimates of female speakers. As the correlation between confidence and accuracy was weak, it was concluded that confidence should not be relied on as an indicator of accuracy in estimation of speaker age.

The three studies in this thesis provide some insight into different aspects of perception of speaker age. Possible implications of the results and suggestions for further research are discussed.

**Keywords:** Age estimation, Voice perception, Speech properties, Speech rate, Vocal disguise, Age disguise, Accuracy, Confidence, Spontaneous speech
Sammanfattning

Syftet med denna avhandling var att undersöka faktorer relaterade till precision vid åldersbedömning av röster, att undersöka vilken betydelse vissa talparametrar har för bedömning och manipulerad ålder, beroende på den typ av talmaterial som använts. Avhandlingen består av tre studier. I studie 1 var syftet att undersöka graden av precision och betydelsen av talhastighet vid åldersbedömning av röster, beroende på lingvistisk variation, dvs. huruvida bedömningen baserades på läst eller spontant tal. I två experiment, ett med läst tal från 36 kvinnliga och manliga talare från tre åldersgrupper (yngre: 20-25 år, medelålders: 40-45 år och äldre: 60-65 år) som stimuli, och ett experiment med spontant tal från samma talare, undersökte vi hur manipulation av talhastighet påverkade lyssnarens åldersbedömningar av unga, medelålders och äldre talare. Resultaten visade att lyssnarna bedömde talarna som yngre vid en ökning av talhastighet med 10% och som äldre vid en minskning av talhastighet med 10%. Effekten av talhastighet var något större i bedömningar av äldre talare jämfört med bedömning av yngre talare, vilket tyder på att talhastighet har större betydelse vid åldersbedömning av äldre röster. Detta mönster var ännu tydligare när lyssnare bedömde ålder baserat på spontant tal. Ökad talhastighet ledde då till att äldre och medelålders (men ej yngre) talare bedömdes som yngre. Överlag bedömdes talare i alla åldersrelaterade veckor vara äldre vid sänkt talhastighet, med undantag för de yngsta talarna. Frånvaron av en linjär effekt av talhastighet i bedömning av spontant tal från yngre talare tyder på att lyssnare använder olika strategier för åldersbedömning beroende på talares ålder och talets spontanitet.

I studie 2 undersökt förändringar i två åldersrelaterade röstparametrar (grundton och talhastighet) när talare försöker ändra rösten i avsikt att låta yngre respektive äldre än de är, och huruvida manipulationerna motsvarade faktiska åldersrelaterade förändringar i grundton (F0) och talhastighet. Studien syftade också till att avgöra hur framgångsrik förståelse av rösten, i syfte att påverka uppfattad ålder, är. Deltagare från tre åldersgrupper (20–25, 40–45 och 60–65 år) läste en kort text under tre röstbetingelser. Tolv talare ingick i varje åldersgrupp (sex kvinnor och sex män). De använde sin naturliga röst i en betingelse och försökte låta 20 år yngre i en andra betingelse och 20 år äldre i en tredje betingelse. Sextio lyssnare bedömde ålder baserat på röstprover från de tre röstbetingelserna. Resultaten visade att talarna höjde F0 och talhastighet när de försökte låta yngre och sänkte F0 och talhastighet när de försökte låta äldre. Röstmanipulationerna påverkade
åldersbedömningarna i riktning med förställningen, men medelvärdet för effekten var endast 3 år. Dessutom grundade lyssnare ålderbedömningen på talhastighet, men inte på F0. Slutsatsen drogs att talare har förmåga att påverka hur andra uppfattar deras ålder genom förställning av rösten, även om den upplevda effekten var mindre än avsett.


De tre studier som ingår i denna avhandling ger inblick i olika aspekter av åldersbedömning av röster. Möjliga implikationer av resultaten och förslag till vidare forskning diskuteras.
List of papers


Introduction

Of all sounds in our auditory environment, the human voice is the most important one. The voice has even been compared to an auditory face, allowing us to not only recognize and discriminate between individuals, but to also draw conclusions about their personal attributes, intentions and emotions (Belin, Fecteau, & Bedard, 2004). It has been suggested that humans are born with an innate neurological predisposition to process vocal information connected to identity and voice recognition (Beauchemin et al., 2011). Thus, the voice is more than a carrier of language. The paralinguistic variation in speech is perceptually distinguished from linguistic information but has the capacity to complement and alter the meaning of what is said to some extent (Belin et al., 2004). Alongside with information about emotion and motivation, speech tells us something about basic personal characteristics. When we hear unfamiliar speakers in everyday life, speech cues to attributes such as gender, age and social background are typically extracted more or less implicitly and automatically (Coleman, 1976; Podesva & Callier, 2015), while explicit assessments of personal characteristics may be required in certain situations. In crime investigations for instance, ear witnesses are often asked to provide information about speaker characteristics. In addition, witness testimonies often contain assessments of the age, gender, accent and dialect of unknown perpetrators (Fashing, Ask & Granhag, 2004). Witness testimonies does indeed play an important role in crime investigations (Kebbell & Milne, 1998; Levi, 1998), but to evaluate the reliability in offender descriptions based on speech, more knowledge is needed regarding how different factors relate to, affect and sometimes challenge their accuracy.

Voice perception

Speech communication is unique to the human kind and is a primary signal in our auditory environment. Most of us pay particular attention to speech sounds and spend more time listening to speech than to any other sounds. Considering the central place that spoken speech has in our everyday life, it is easy to view the voice merely as a transporter of language. However, from an evolutionary perspective, speech is a recent phenomenon that appeared quite late in human evolution, while other types of vocal communication has been of central importance to our ancestors for millions of years (Fitch, 2000). Long before speech developed, interpretation of vocalisation was important for the survival of our ancestors. For instance, recognising vocalisation from predators and prey as well as the voices of family members and
socioemotional content in their vocalisation were crucial abilities (Fitch). Hence, the human evolution has had a lot of time to refine abilities to infer, for example, emotional content from vocal cues. Evidently, listeners are able to accurately decode vocally expressed emotions (Juslin & Laukka, 2001; Juslin & Laukka 2003), and perceive spontaneously expressed emotions as more genuine compared to emotional expressions posed by actors (Juslin, Laukka & Bänziger, 2018). Furthermore, processing of emotional expressions is primarily implicit (Niedenthal & Showers, 1991) and it has been argued that humans have an innate neurological preparedness to process voices and to recognise speakers (Beauchemin et al., 2011). Thus, when we hear someone speak, we automatically make a variety of inferences about the situation and the person speaking, almost as if we could hear the speaker’s personal attributes and emotions (Krauss, Freyberg & Morsella, 2002). However, the fact that we implicitly perceive social attributes based on speech does not mean that the inferences are always accurate. Voice identification is clearly less accurate than face recognition (Cook & Wilding, 2001; Stevenage, Howland & Tippelt, 2009) and assessments of features such as speaker age (Braun, 1996; Braun & Cerrato, 1999; Brückl & Sendlmeier, 2003; Hollien and Tolhurst, 1978; Huntley et al., 1987) and accent (Brunner, 2009) have been found to be biased.

Fujisaki (2004) have categorised information conveyed in speech into three different types; linguistic, paralinguistic and non-linguistic. Linguistic information is described as symbolic information containing lexical, syntactic, pragmatic and semantic information (e.g. words, phrase structure and discourse). Paralinguistic information is defined as information that cannot be inferred in written words, but that is inferable in speech, and that has the potential to modify or add meaning to the linguistic information (e.g. intentions and attitudes). According to Fujisaki, the paralinguistic information can be modified deliberately by the speaker. The non-linguistic information, on the other hand, is claimed to be hard to control deliberately and includes information about the age, gender and emotional state of the speaker. However, Fujisaki stresses that the categorisation is not strict and that the boundaries between the different types of information is not always clear. For instance, it is not obvious that attitudes and intentions are always expressed deliberately, while emotions are not. In this thesis, I will not differentiate between paralinguistic and non-linguistic information. My use of the term paralinguistic information will thus also refer to what Fujisaki describes as non-linguistic information, while the term linguistic information will be used according to his categorisation.
Since vocalisation conveys information about the same aspects as faces do, the voice has been described as an “auditory face”. Belin et al. (2004) proposed a model of voice perception based on Bruce and Young’s information processing model of face perception (Bruce & Young, 1986). Bruce and Young’s theoretical framework has contributed significantly to understanding of facial information processing. They suggested that different types of information code, functionally separate, could be extracted from faces (e.g. identity-specific, semantic, expression and facial speech codes). According to Belin et al (2004), the process of voice perception resembles that of face perception in important ways. Belin et al. divide vocal information processing into three systems that are functionally independent; analysis of speech information (linguistics), analysis of affective information and analysis of identity information. Of course, vocal signals usually contain more than one type of information at the same time. For example, when we hear someone speak, we may hear what that person is saying, what mood s/he is in and we may pick up cues to personal characteristics or identity. Correspondingly, neuroimaging studies have demonstrated that linguistic, affective and identity information is processed in partially separated pathways in the brain (Imaizumi et al., 1997; Nakamura et al., 2001; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003) There are also clinical cases where patients have demonstrated preserved speech perception while speaker recognition has been impaired (Belin et al., 2004). Yet, there is potential for integration between vocal and facial information processing. For example, lip-reading can support speech comprehension and facial features are important for qualities in sign language. Integration between facial and vocal information processing has been found for linguistic content in speech (Romanski, 2012), affective information (Collignon et al., 2008; De Silva, Miyasato, & Nakatsu, 1997; Park et al, 2010) as well as in identity processing (Campanella & Belin, 2007). Nevertheless, it is evident that face processing (Herzmann, Schweinberger, Sommer & Jentzsch, 2004) as well as voice processing (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) are highly specialised processes.

To summarise, the perception of speaker characteristics and identity can be described as specific processes important in person perception. The basis for perception of identity relevant information are regularities in the variation of vocal features related to personal characteristics (e.g. age and gender) and identity (speaker recognition). For example, the unique combinations of invariant vocal properties within a speaker’s voice allow for recognition of that speaker, while variation between speakers in vocal properties, such as
speech rate and fundamental frequency, allow for age perception. Regarding age perception, the perceptual correlates to speaker age corresponds, to some extent, to actual changes in voice and speech caused by ageing.

Age related changes in speech production
The human voice changes from early childhood and throughout life. The vocal cues that makes estimation of speaker age possible, derives from anatomical and physiological age related changes in the speech production mechanism. The phonatory organ can be divided into three components; the compressor (the lungs), the resonator (the larynx) and the articulator (supralaryngeal structures). The supralaryngeal structures filter the sound produced at the resonator and include the vocal tract (the laryngeal cavity, the pharynx, as well as the nasal and oral cavities), and oral muscles (tongue, lips, jaw muscles). When changes occur in the structures of the phonatory organ, it has impact on speech performance and voice quality. But the vocal changes do not alone account for the all age information conveyed in speech. Age related changes in cognitive functioning also play an important role as it has impact on linguistic performance, complexity and speed. Thus, age can be perceived both through voice quality, speaking tempo and in aspects of language production.

The vocal ageing process is intertwined with hormonal changes and evolves in different ways and with different timing depending on speaker sex (Bürckl & Sendlmeier, 2003; Linville, 2001; Martin, 1997). It should be noted that individual differences in vocal ageing are substantial (Keller, 2006; Ramig & Ringel, 1983) and that environmental factors such as smoking, and disease-related changes such as vocal fold edema, can have effects on speech similar to those that depend on ageing. In many cases it is hard to distinguish between vocal changes caused by normal ageing and changes caused by environment or pathology. Yet, studies of age related changes in speech production have found some general patterns in the ageing of the speech production mechanism, which are perceived by listeners as alterations mainly in voice quality, pitch and speech rate.

Speech production rely on various systems that interact in complex ways. This section does not claim to give a full review of all age related changes in speech production, yet it describes a selection of important physiological and anatomical changes with effects on speech production, which are likely to be meaningful to perception of speaker age.

It is not always obvious exactly how certain age related changes in physiology, anatomy or cognition are related to perceivable changes in speech
production with increased age. In this section, important changes in speech production related to age are presented first. Thereafter, physiological and anatomical changes related to age and speech production are presented, in some cases with suggested effects on speech properties.

**Changes in speech properties**

Speech production is affected in several ways by the ageing process, which is reflected in measurable differences between younger and older speakers in numerous speech properties. Pitch is a distinct vocal parameter that plays a particularly important role in speech communication (Oxenham, 2012). Modulations of vocal pitch are closely tied to prosody as well as to emotional expression (Breitenstein, Lancker, & Daum, 2001; Bänziger, & Scherer, 2005; Juslin & Laukka, 2001), and have the potential to affect the meaning of spoken words (Oxenham, 2012). Vocal pitch depends on fundamental frequency (F0) to a large extent (Kawahara, Masuda-Katsuse, De Cheveigne, 1999; Klatt, 1973), but also on variation in the amplitudes of formant frequencies, or overtones, which correspond to certain resonances in the vocal tract and define the quality of different vowels. While F0 is produced at the larynx, the formants are shaped in the vocal tract. In women, F0 decreases with 30 Hz in general during adulthood, and most profoundly after menopause, while in men, the decrease in F0 during adulthood is smaller, followed by a pronounced increase by approximately 30 Hz in late adulthood (Linville, 2001). The pattern reported by Linville (2001), has been replicated by Dehqan, Scherer, Dashti, Ansari-Moghaddam, & Fanai (2012). Thus, there is no general linear decline in F0 with increased age. Instead, the patterns of age related changes in F0 are rather complex, and gender dependent (Brückl & Sendlmeier, 2003; Dehqan et al., 2012; Linville, 2001; Martin, 1997). The same goes for formant frequencies, for which several studies have found changes related to increased age (Harrington, Palethorpe, & Watson, 2007; Hawkins & Midgley, 2005; Torre & Barlow, 2009). For example, Hawkins and Midgley (2005) reported that F1 was lower in /ε/ and /æ/, and F2 to be lower in /u:/ and /ʊ/, in older speakers compared to younger ones. In a longitudinal study by Harrington et al. (2005) speakers displayed lower F1 and F2 in later recordings compared to earlier ones.

Maximum vocal intensity, corresponding to perceived loudness and vocal effort, has been found to decrease with increased age (Hodge, Colton & Kelley, 2001; Linville, 2001). It should be noted however, that in conversational speech, speakers are seldom required to use their maximum vocal intensity, why this change might not be distinguishable in everyday speech.
During late adulthood, voices of both women and men become more unstable, with increased levels of jitter and shimmer and lower harmonics-to-noise (H/N) ratios (Decoster & Debruyne, 1997a; Dehqan et al., 2012), as well as increased F0 variability (Linville, 2001). Jitter and shimmer have been reported to correlate with perceived vocal roughness and breathiness (Heiberger & Horii, 1982; Kreiman & Gerratt, 2000), and to contribute to overall patterns of voice quality (Kreiman & Gerratt, 2005). The H/N ratio reflects the degree to which noise replaces the harmonic structure in a sustained vowel, and has been described as a psychophysical measure of the degree of hoarseness in voice (Yumoto, Sasaki, & Okamura, 1984). Thus, vocal hoarseness and roughness is related to speaker age, such that greater vocal roughness and hoarseness, corresponding to higher jitter and shimmer and lower H/N ratio, is observed in older speakers than in younger speakers (Decoster & Debruyne, 1997a; Dehqan et al., 2012).

A particularly clear relationship has been observed between speech rate and speaker age. Speech rate can be measured in various ways. A few examples of measures are syllables per second, segment duration, syllable duration and sentence duration. Evidence of decreased speech rate with increased age has been presented in several studies (Bürckl & Sendlmeier, 2003; Linville, 2001; Oyer & Deal, 1985; Smith, Wasowicz, & Preston, 1987). For example, Linville (2001) reported 20-25% longer durations in syllables and sentences, as well as greater speech rate variability, in older speakers compared to young adult speakers. Age related changes in voice-onset time has also been reported in several studies (Decoster & Debruyne, 1997b; Neiman, Klich, & Shuey, 1983; Sweeting & Baken, 1982).

Changes in speech production systems

The vocal changes described above are based on changes in several systems that take part in speech production. In this section, a selection of important changes in speech production systems is described.

The larynx is a complex structure that controls age related vocal features such as pitch and intensity. The larynx descends, relative to the vertebrae, and increases in size during childhood, reaching its full size in puberty. However, after puberty the larynx continues to change, affecting fundamental frequency, voice quality and intensity (Linville, 2001). Atrophy takes place in the intrinsic phonatory muscles in the larynx, as well as in the vocal folds, in both men and women in late adulthood (Hirano, Kurita & Sakaguchi, 1989; Linville, 2000; Martins et al., 2015) which leads to decreases in vocal power and intensity (Martins et al.). Due to hormonal and neurological tissue alterations that occur
with age, the structure and length of the vocal folds are affected (Linville, 2000). The vocal folds shorten in men, especially in older age (Hirano, Kurita & Sakaguchi, 1989). Further, the thin layer of tissue (the epithelium) that protects structures in the larynx, thickens rather continuously with ageing in women and men (Hirano et al., 1989). But while the thickening of the epithelium increases during late adulthood in women, it grows thinner in men during the same age period (Hirano et al.). In both women and men, the vocal folds loses thickness with increased age, which may lead to an inability in the vocal folds to close fully during vocalisation (Tarafder, Datta, & Tariq, 2012). Linville (2002) report greater variability in spectral amplitudes in older men, and describes the findings as acoustic support for the increased occurrence of glottal gaps with increased age.

Furthermore, the secretion in the mucous glands, which hydrate the vocal folds, is reduced in older speakers (Sato & Hirano, 1998). The dryness of the vocal folds may lead to instability in the vibration, particularly in older men (Tarafder, Datta, & Tariq, 2012). Moreover, Linville (2001) has described loss of elasticity and strength in the respiratory muscles as well as stiffening of the thorax, leading to decreased lung capacity, with increased age. Linville further reports that the facial muscles lose elasticity due to ageing, with potential effects on articulation. She has also reported a 3-5% growth in the craniofacial skeleton during adulthood, which may have effects on formant frequencies. Decreases in vowel formant frequencies associated with enlargement of the vocal tract has been reported by Endres, Bambach and Flosser (1971), and by Linville and Fisher (1985). Linville and Fisher observed age related changes in the vocal tract and found a decrease in the first formant frequency (F1) in the vowel /ae/. Others have found increases in F1 levels in front and mid vowels with older age, while back vowels were significantly lower (Rastatter & Jacques, 1990), and have suggested that the changes in formant frequencies may appear due to alterations in articulatory positioning, rather than enlargement of the pharynx. Linville (2001) however, presents a “mixed model”, to explain resonance effects, which takes account of age and gender interactions, and proposes that changes in formant frequencies are effects of both laryngeal lowering and articulatory patterns. It is clear that the changes in formant frequencies, that might be dependent on age related changes in the vocal tract, cannot be described in terms of a general decline (Xue, Jiang, Lin, Glassenberg, & Mueller, 1999).

With ageing, neuromuscular function decreases occur, in the peripheral as well as the central nervous system, affecting speech production. For example, sensorimotor processes are affected by age related reduction of dopamine
levels and neurons in the cortex. The decreased neuromuscular control, due to nerve degeneration and hormonal changes, impacts functioning in laryngeal muscles as well as in oral muscles, such as the tongue, lips and jaw (Linville, 2001). Related to the changes in neuromuscular processes, a decline in articulation rate has been observed (Linville, 2001). The age related decrease in speech rate during adulthood is probably also caused by cognitive factors, especially in older speakers. Burke and Shafto (2004) reported deficits in linguistic production with increased age, marked by reformulations, difficulty in finding the intended words and an increase in number of pauses. However, while word retrieval during speech, was observed to decline with older age, there was also a parallel pattern of increased linguistic complexity and semantic knowledge with increased age (Burke & Shafto). The advantage observed in older speakers, in the ability to produce long and complex utterances, tends to appear when situations are less controlled. In controlled situation, where working memory capacity is challenged, older speakers instead tend to produce less complex utterances compared to younger speakers (Kemper, Thompson, & Marquis, 2001; Kemper, Herman & Lian, 2003). Furthermore, age related linguistic changes have been reported regarding the use of emotion words, such that older speakers tend to use more positive emotion words and fewer negative ones than younger speakers do (Pennebaker & Stone, 2003). Thus, ageing impairs many speech functions while affecting others in more positive ways.

Perception of speaker age

Since age is one of the most important social factors, we use our ability to assess the age of other people in many situations in everyday life (Moyse, 2014). Age related information is clearly important in person perception. For example, age factors are related to perception of personality (Etaugh & Birdoes, 1991), credibility (Masip, Garrido, & Herrero, 2004) and effectiveness (O’connell & Rotter, 1979). Based on perceptions of the age of other people, we regulate our behavior relative to the social traits of those we are socialising with. It has, for example, been demonstrated that people adjust the way they speak distinctly, depending on the age of the person they are talking to (Hummert, Shaner, Garstka, & Henry, 1998; Kemper, Finter-Urczyk, Ferrell, Harden, & Billington, 1998).

It turns out that age estimation is a skill that most people master pretty well. Strong correlations (0.70-0.90) between chronological speaker age and estimated speaker age have been reported in several studies (e.g. Braun, 1996; Braun & Cerrato 1999; Brückl & Sendlmeier, 2003; Gnevsheva & Bürkle, 2019;
Huntley, Hollien & Shipp, 1987; Krauss et al. 2002; Shipp & Hollien, 1969). The accuracy (absolute mean deviation from chronological age) in estimation of speaker age has been reported to be 7-11 years (e.g. Hughes & Rhodes, 2010; Krauss, Freyberg & Morsella, 2002; Schötz, 2006), which can be compared with the accuracy in age estimation of faces that has been found to be 4-7 years (e.g. Dehon & Brédart, 2001; Krauss, Freyberg & Morsella, 2002; Sörqvist & Eriksson, 2007; Vestlund, Langeborg, Sörqvist & Eriksson, 2009; Voelke, Ebner, Lindenberger & Riediger, 2012). However, several factors affect the level of accuracy in age estimates. For instance, accuracy is influenced by the personal characteristics of both speakers and listeners, by the length and cue-richness in the speech stimuli, and of course, is dependent on how precise an estimate is demanded to be for it to be considered accurate. Yet, the methods used in studies on estimation of speaker age vary, and it can therefore be hard to compare results from different studies. For example, effects found in experiments using highly controlled stimuli, such as sustained vowels or isolated words, might not be valid for the spontaneous conversational speech, rich in different types of speech cues, that we are exposed to in everyday life.

**Perceptual cues to speaker age**

Perceptual cues to speaker age are speech properties that listeners rely on when they estimate age. Even if perceptual cues to speaker age are likely to be based on actual age related changes in speech, they do not need to be equivalent to actual age related changes.

The amount of research performed on perceptual cues to age is much smaller than that about vocal ageing. Thus, it is not yet clear which properties are the most important ones, but several studies have found that age estimates are affected by speech rate (Harnsberger, Shrivastav, Brown Jr, Rothman, & Hollien, 2008; Schötz, 2004; Schötz, 2006; Shipp, Qi, Huntley, & Hollien, 1992; Stölten & Engstrand, 2003; Winkler 2007). Breath management, measured as numbers of breaths and breath duration, is another temporal cue that has been suggested to impact age estimation (Shipp et al., 1992). Furthermore, fundamental frequency has been found to be important in some studies (Schötz, 2006; Shipp et al., 1992; Stölten & Engstrand, 2003) but not in other ones (Harnsberger et al., 2008). Yet, Schötz (2006) has reported that formant frequencies might be more important than both speech rate and F0. She has also demonstrated gender differences in perceptual cues and indicated that F0 and speech rate seem to be more important in estimates of female speakers age than in estimates of male speakers age (Schötz, 2004). Brückl and Sendlmeier (2003) used sustained vowels, as one of other speech materials,
and found that age estimation accuracy increased when the vowel onset was included in the speech sample.

**Measuring age estimation**

To a large extent, age estimation accuracy is task-dependent. In some studies on estimation of speaker age, the listeners have been instructed to merely categorize the speakers in to broad age ranges (e.g. Neiman & Applegate, 1990; Cerrato, Falcone & Paoloni, 2000; Ptacek & Sander, 1966; Jacques and Rastetter, 1990). This type of task is suitable for studies where the listeners are children (Anastasi & Rhodes, 2005). One disadvantage in using age categories (e.g. “50-60 years”) rather than exact age estimates (e.g. “53 years”) is that it is a relatively imprecise measure of accuracy. It might also overestimate listeners’ accuracy. For example, in a study by Ptacek and Sander (1966) the listeners made correct age estimations in 99% of the cases when they estimated age based on read speech. Another issue is the statistical methods applied when age categories. In the studies by Ptacek and Sander (1966) as well as Jacques and Rastatter (1990) accuracy was calculated as the percentage of correctly identified age categories. Yet, the percentages will be dependent on the number of age categories in the experimental task. Neiman and Applegate (1990) and Cerrato et al. (2000), who also used age categories, calculated the accuracy as the correspondence between estimated and chronological age. Correspondence was calculated by dividing the lower value with the higher one, which allows for greater differences between estimated and chronological age in estimates of older speakers’ age than in that of younger speakers’ age.

It is often more useful to ask listeners to make exact age estimates and to employ statistical methods that allow for comparison between different studies. Yet, in studies using exact age estimates, the correlation between estimated age and chronological age is frequently used as a measure of accuracy (e.g. Braun & Cerrato 1999; Cerrato et al. 2000; Gocsal, 2018; Hartman, 1979; Huntley et al. 1987; Krauss et al. 2002; Neiman & Applegate, 1990; Ryan & Burk, 1974; Shipp & Hollien 1969). In these studies, high correlations are interpreted as high levels of accuracy. Thus, it means that accuracy is defined as a linear relationship between estimated and chronological age, without respect to systematic under- or overestimations. The most plausible measure, in terms of exactness and usefulness, is the absolute mean deviation, calculated as the unsigned mean difference between estimated and chronological age. Absolute mean deviation has been used in some studies on estimation of speaker age (e.g. Hughes & Rhodes, 2010; Krauss, Freyberg &
Morsella, 2002; Nagao & Kewley-Port, 2005; Schötz, 2006) and provides an exact value of deviation from chronological age. Unsigned values are used to prevent over- and underestimation to cancel each other out. However, the signed mean deviation can be used to study systematical over- and underestimations of chronological age.

The influence of speaker characteristics and speech material on age estimation performance

It has been suggested that listeners use different age estimation strategies depending on the gender of the speaker, as age related changes in speech properties are highly gender dependent (Schötz, 2005; Schötz 2006). Regarding differences in accuracy, inconsistent results have been presented. While some studies have found no difference depending on gender (Hughes & Rhodes, 2010; Krauss et al., 2002; Mulac & Gilles, 1996), others have found that the age of female speakers is estimated with higher accuracy than that of male speakers (Harnsberger, Brown, Shirastav, & Rothman, 2010; Neiman & Applegate, 1990). Speaker gender interacted with age in Harnsberger et al. (2010) and higher accuracy was found in estimates of female speakers in the youngest age group (18-30 years), while no gender differences in accuracy was found for middle aged (40-55 years) or older speakers (62-92 years). The results from these studies suggest that there might be an advantage in accuracy for estimates of the age of female speakers, and particularly young females.

When it comes to the chronological age of the speakers, it is quite clear that accuracy is higher in estimates based on younger adult speakers rather than on older adult speakers (e.g. Huntley et al., 1987; Jacques & Rastatter, 1990; Moyse et al., 2014). A bias towards the population mean has been found in several studies and seems to be a particularly robust effect in estimation of speaker age. While the age of older speakers is systematically underestimated, the age of younger speakers is systematically overestimated (Braun, 1996; Braun & Cerrato, 1999; Brückl & Sendlmeier, 2003; Hollien and Tolhurst, 1978; Huntley et al., 1987).

The level of difficulty in age estimation tasks vary depending on how rich the speech material is, in perceptual cues to speaker age. Most previous studies on estimation of speaker age have used read speech, isolated words or vowels as stimuli (Braun, 1996; Braun and Cerrato, 1999; Cerrato, Falcone, & Paoloni, 2000; Harnsberger et al., 2008; Huntley et al., 1987; Ptacek and Sander, 1966; Ramig and Ringel, 1983; Shipp, Qi, Huntley & Hollien, 1992; Torre and Barlow, 2009; Winkler, 2007), while spontaneous speech is less
common. In a study by Brückl & Sendlmeier (2003) estimates based on sustained vowels were found to be less accurate than estimates based on read speech, and read speech was found to generate less accurate estimates than spontaneous speech did. Naturally, passages of spontaneous speech are richer in different types of age related information than sustained vowels and read speech are. While sustained vowels only contain vocal cues, read speech also contain prosodic and temporal cues and spontaneous speech further adds cues in terms of linguistic variation. The argument for using read speech rather than spontaneous speech is that it allows for control over the linguistic content in the speech material. Accordingly, speech properties that have been found to be important cues to speaker age in experiments were all speakers read the same text, might not be as relevant to daily life, where age perception is mainly based on spontaneous speech. Further, it is possible that listeners use different strategies and rely on different cues, depending on the type speech material, as suggested by results in Schötz (2006) and Brückl & Sendlmeier (2003). Hence, it is meaningful to investigate if certain speech properties that has been found to be important cues in read speech, are used by listeners to the same extent when age estimates are based on spontaneous speech. After all, spontaneous speech is more common than read speech in real life situations where age estimation is performed.

Age estimation in ear witness testimony

Through the history of witness psychological research, a visual culture has dominated the field, and relatively little attention has been dedicated to ear witness testimony. One explanation for the asymmetry might be that accuracy in identification and estimation of personal characteristics is generally lower when the testimony is based on auditory stimuli rather than on visual stimuli (McAllister, Dale, Bregman, McCabe & Cotton, 1993; Yarmey, 1995). For example, it is easier to correctly identify an unfamiliar face than an unfamiliar voice, and age estimates of faces are more accurate than age estimates of voices (McAllister et al., 1993). It has also been demonstrated that people in general seem to automatically prefer processing of face information before speech in order to facilitate later recognition (Cook and Wilding, 2001; Stevenage, Howland & Tippelt, 2009; Tomlin, Stevenage, & Hammond, 2017), while speech information is primarily processed semantically when presented in combination with a face (Cook and Wilding, 2001), an effect known as the Face Overshadowing Effect. However, visual conditions are seldom optimal at actual crime scenes. Vocal information might be the most salient cue to a perpetrator’s identity and personal characteristics when the perpetrator has
been observed in darkness, been masked, heard over the phone, or when the victim has been purposely hindered from seeing the perpetrator. To many people, who are blind or in other ways visually impaired, auditory perception is the primary sense used in order to recognise and socially categorise others.

Research on both ear and eye witness testimony, has too a large extent focused on identification (e.g. Yarmey, 1995) while offender descriptions and assessments of personal characteristics has been somewhat overlooked. Thus, while offender descriptions play an important role in crime investigations, knowledge about accuracy in witness testimonies is largely based on studies of identification. One well-studied topic is the relationship between witnesses’ confidence that their identifications are correct and the actual identification accuracy. The relationship has been found to be weak, in eyewitness identification (e.g. Sporer, Penrod, Read & Cutler, 1995) as well as in earwitness identification (e.g. Zetterholm, Sarwar, Thorvaldsson, & Allwood, 2012). In other words, an identification made by a witness who is confident that s/he is correct, should not be expected to be more accurate than an identification made by someone who is less confident. The correlation between confidence and accuracy has however been found to be stronger when visual conditions are optimal (Lindsay, Read & Sharma, 1998), which is not always the case at actual crime scenes. Anyhow, investigators and courts still frequently ask witnesses to rate how confident they are that their testimonies are correct. The confidence-accuracy relationship in ratings regarding identifications of faces is well-studied (Brewer, Keast, & Rishworth, 2002; Juslin, Olsson, & Winman, 1996; Palmer, Brewer, Weber, & Nagesh, 2013). Yet, from a witness psychological point of view, it is also of interest to examine if confidence can be an indicator of accuracy in descriptions of personal characteristics such as age, based on vocal information. For in cases where there is no suspect, identification is not possible. In either case, it is crucial to gather information about the perpetrators characteristics to narrow down the options of possible suspects. Witnesses play a key role in such processes (Cutler, Penrod & Dexter, 1990) and age is one of the most frequent characteristics included by witnesses in descriptions of unknown perpetrators (Fahsing, Ask & Granhag, 2004). Of course, age assessments are likely to be influenced by a multitude of factors. Apart from factors that have been observed to affect age estimation in experimental studies, witnesses of actual crime are likely to be subject to emotional stress and various distractions. Speaker identification and offender descriptions may sometimes be further afflicted by the fact that criminals occasionally use voice disguise in order to obstruct identification (Boë, 2000; Neuhauser, 2008; Orchard and Yarmey,
According to Perrot and Chollet (2012) voice disguise is “applied when there is a deliberate will to transform one’s voice to imitate someone or just to change the sound”. An archive study by Künnzel (2002) at the speaker identification section at BKA (The German Federal Police) reported that 15-25% of the cases processed contained some kind of voice disguise. Some of the disguise techniques were performed using electronic devices while other were non-instrumental, such as whisper, falsetto, use of quirky voice, imitation of dialect or foreign accent as well as modulation of the voice in order to sound another age. The technique used by a perpetrator can tell us something about the intention behind the disguise. While someone who uses whisper probably do so to hide his or her identity, age disguise has a more specific purpose. In cases of online grooming, for instance, perpetrators often pretend to be younger than they are in order to attract children (Ashcroft, Kaati & Meyer, 2015). Thus, knowledge about age disguise is relevant to fight the growing amount of cyber related crime, particularly in cases of online grooming. One important step in the development of methods to discover age disguise is to gain knowledge about what adults do in an effort to be perceived as younger, and about the degree to which the disguise is successful. Studies using linguistic text analysis have found that adults pretending to be children can be separated from actual children based on writing style (Ashcroft et al, 2015; Chiang & Grant, 2017). Once the groomer has built a relationship with the child online, telephone contact may be initiated requiring not only linguistic but also vocal age disguise. In these cases speech is modulated to sound younger. Due to age related changes in the speech production mechanism it is possible that older speakers might be less able to manipulate certain vocal aspects, and features related to their actual age might “leak” through the attempted disguise. Therefore, it would be useful to identify indicators of age disguise, to gain knowledge about how successful age disguise is and to identify voice properties that are resistant to age disguise.

In a previous study about vocal age disguise by Lass et al, (1982), the speakers were instructed to modulate their speech in order to sound younger in one condition and older in another condition. Only small differences were found in perceived age, in the attempted direction. No analyses was performed on changes in speech properties related to age disguise and thus, the study provided no information on how the speakers changed their voices and speech in order to sound younger and older. In a recent study by Hautamäki, Kanervisto, Hautamäki and Kinnunen (2018) listeners received information that some voices were modulated with intention to sound older
or younger and the listeners were asked to estimate both chronological and intended age, by classifying the speakers into broad age categories. In some cases, the age disguise was relatively extreme. For instance, adult voices were modified to sound like children’s voices and to sound older than 81 years. The results indicated that listeners’ estimations of the intended age of the disguised voices followed the direction of the disguise in general. Thus, listeners seemed to be able to assess the direction of the disguise to some extent. However, the study by Hautamäki et al. (2018) does not tell us much about how speakers modulate their voices to disguise age, nor whether age disguise is successful when listeners do not know that the speakers might be disguising their voices to sound younger or older.

In sum, research on age estimation clearly has an important applied value from a forensic psychological perspective. However, most studies on age perception have not been performed to investigate age estimation as an aspect of ear witness testimony. It is therefore important to design age estimation studies considering witness testimony as an applied context.
Aim

The aim of this thesis is to investigate factors related to accuracy in estimation of speaker age, as well as the role of certain speech properties in perception and manipulation of speaker age, with respect to the material that the age estimates are based on.

Most studies on estimation of speaker age have used read speech as stimuli. However, in everyday life, as well as in cases of ear witness testimony, estimates are often based on spontaneous and/or conversational speech. It is not yet clear if the speech properties that have been found to be important cues to speaker age in studies using read speech, are as important when the estimates are based on spontaneous speech, which contain linguistic cues to age that are absent in read speech. Through manipulation of certain speech properties, effects on perceived age can be studied directly, relative to speech material. The speakers themselves can also perform manipulation, and it is of interest whether speakers are able to modulate their voices in order to sound younger or older, and to study how speech properties are affected by such attempts. From a forensic perspective, it is also important to identify indicators of accuracy in estimation of speaker age. Confidence is often used as an indicator of accuracy in witness testimony although the relationship between confidence and accuracy has been found to be weak. The studies in this thesis were carried out with the aim to clarify some of the concerns raised here related to accuracy, speech properties and speech material.

In Study 1 the aim was to investigate the role of speech rate as well as the level of accuracy in estimates of speaker age depending on linguistic variation in the speech material. In Study 2 the aim was to investigate how speech rate and fundamental frequency are affected when speakers disguise their voices in order to sound younger and older, and whether age disguise had an effect on estimated age. In Study 3 the aim was to study confidence and accuracy in estimation of speaker age and whether confidence can serve as an indicator of estimation accuracy.
Summary of studies

Ethical considerations
The studies in this thesis were conducted in accordance with the Declaration of Helsinki and the ethical guidelines given by the American Psychological Association. All participants (listeners and speakers) were adults and gave their informed consent. Both listeners and speakers signed an information agreement form. The experiment caused no harm, the identities of the participants were kept confidential, and no conflicts of interest were identified.

Study 1

Background and aim
The human voice goes through many age related changes throughout life. Physiological and anatomical changes occur with increased age and have effects on for example speech rate, fundamental frequency and voice stability. Thus, listeners may rely on several cues when they infer age from speech. Speech rate is a voice property that has been found to be a particularly important cue to speaker age (e.g. Harnsberger et al., 2008; Schötz 2004; Winkler, 2007). However, most studies on estimation of speaker age have been based on read speech (e.g. Braun, 1996; Braun & Cerrato, 1999; Cerrato et al., 2000; Harnsberger et al., 2008; Huntley et al., 1987; Ptacek & Sander, 1966; Ramig & Ringel, 1983; Shipp et al., 1992; Torre & Barlow, 2009). Yet, in everyday life, speech is often conversational and comes about spontaneously. As spontaneous speech not only contain non-linguistic cues to speaker age, but also linguistic ones, it is likely that listeners would make more accurate estimates based on spontaneous speech than on read speech. Listeners might also rely less on non-linguistic cues, such as speech rate, when age is estimated based on spontaneous rather than read speech. Furthermore, only a few studies hitherto (Harnsberger et al., 2008; Schötz, 2004; Stölten & Engstrand, 2003; Winkler, 2007) have investigated the effects of speech rate on perceived age by actually manipulating speech rate. Among these, the study by Harnsberger is the most relevant one, as they used stimuli longer than a few words. However, Harnsberger et al. did not study the effects of increased speech rate on perception of young speakers, nor the effects of decreased speech rate on perception of older speakers. Moreover, Harnsberger et al. manipulated the speech rate with plus and minus 20%, which is a quite
substantial manipulation. A preliminary study performed before Study 1 in this thesis indicated that a manipulation of such magnitude made it obvious to the listeners that some of the speech samples were manipulated, while the speech samples were perceived as natural when a manipulation of plus and minus 10% was used. We therefore decided to use a more subtle manipulation of plus and minus 10% to study the effects of speech rate on perceived age. In two experiments, one using read speech as stimuli and the other using spontaneous speech, we investigated how changes in speech rate influenced listeners’ age estimates of young adult, middle aged and older speakers. We hypothesised that decreased speech rate would generate higher age estimates while increased speech rate would generate lower age estimates. We also explored whether the magnitude of the speech rate effects depend on the chronological age of the speakers. Furthermore, it was tested whether the impact of speech rate would be smaller, and if accuracy would be higher, when spontaneous speech was used, rather than read speech.

**Procedures, design and analyses**

Eighty-one listeners estimated the age of 36 speakers from the age groups 20-25 (young), 40-45 (middle aged), and 60-65 years (older). There were 12 speakers from each age group. In Experiment 1 the speakers were recorded while reading a 35-word text containing walking directions, while in Experiment 2, the speakers were asked to provide directions on how to navigate from an origin to a destination on a map. The speech rate in the speech samples was edited in the same way in Experiment 1 as in Experiment 2. Three versions of each recording were used in the listening tests; one un-manipulated, one with the speech rate increased by 10%, and one with the speech rate decreased by 10%. Fundamental frequency for each speech sample was analysed in Praat and it was found that there was no difference in F0 depending on age group. Thus, F0 was not included as a factor in subsequent analyses.

The listening tests were conducted in a laboratory where speech samples were presented to the participants through headphones. The participants were instructed to estimate the age (in years) of each speaker directly after hearing each of them. Each participant estimated the age of each speaker only once. The listeners were randomised into three listener groups that were balanced with regard to gender and age. Each listener group heard 36 speech samples produced by all 36 speakers. Twelve of the voices presented to each listener group was un-manipulated, 12 were decreased in speech rate and 12 were increased in speech rate. Thus, the listeners heard different speakers at
different speech rates depending on listener group. A randomised order was generated for each of the three sets of speech samples.

For both experiments, a 3 (20–25, 40–45, 60–65 years) × 3 (speech rate: increased, natural, decreased) within participants factorial design was used to measure differences in age estimates depending on speaker age group and speech rate. Two dependent measures were calculated; signed differences between age estimates and the chronological age of the speaker (to investigate over-and underestimations) and the absolute/unsigned differences (to investigate accuracy). A cross-experiment analysis on accuracy estimates was also conducted, to test the hypothesis that age estimation accuracy is higher for spontaneous speech than for read speech.

A 3 (speaker age group: 20–25, 40–45, 60–65 years) × 2 (material: read speech, spontaneous speech) repeated measures analysis of variance was performed for estimates of voices at natural speech rate from both experiments. Post hoc analyses were computed using the Bonferroni correction.

**Results and conclusions**

As illustrated in figure 1, listeners overestimated the age of younger speakers and underestimated the age of older speakers. Our hypothesis, expecting that increased speech rate would make the speakers sound younger while decreased speech rate would make speakers sound older, was confirmed by the data (Figure 1). However, the effect of speech rate was dependent on speaker age. The speech rate effect was most pronounced in estimates of older speakers’ age. Although estimates based on increased speech rate was lower than those based on decreased speech rate in all speaker age groups, the difference was larger in estimates of older speakers’ age in comparison with estimates of younger speakers’ age.
Figure 1. Age estimation in Study 1, Experiment 1, calculated as the average of the signed differences between the age estimations and chronological age of the speakers. The estimates were based on read speech, at increased, neutral and decreased speech rate, from young, middle aged and older speakers.

In Experiment 2, the result pattern was quite similar to that found in Experiment 1. Again, there was a main effect of speech rate such that speakers sounded younger when speech rate was increased, and older when the speech rate was decreased (Figure 2). However, an interaction between speech rate and age group revealed that it was only in age estimates of the oldest age group that there was a clear-cut negative relationship between speech rate and age estimates. This interaction reveals that the effect of speech rate is linearly related to age estimates of older speakers—faster speech rate is associated with lower age estimates (i.e., faster speech rate make the speaker sound younger)—but this is not the case in estimates of young speakers—wherein the highest age estimates were found for the natural speech rate. Follow-up t tests showed, in estimates of young speakers, that there was no significant difference between fast and slow speech rate, and no difference between slow and natural speech rate, but there was a difference between fast and natural speech rate in estimates of young speakers. However, for both middle-aged and older speakers, there was a difference between fast and slow speech rate. Taken together, the speech rate effect behaved differently for the
three speaker age groups in Experiment 2. Moreover, both experiments showed that younger speakers were estimated with highest accuracy while older speakers were estimated with lowest accuracy.

Figure 2. Age estimation in Study 1, Experiment 2, calculated as the average of the signed differences between the age estimations and chronological age of the speakers. The estimates were based on spontaneous speech, at increased, neutral and decreased speech rates, from young, middle aged and old speakers.

It was concluded that listeners use speech rate as a perceptual cue to speaker age, but this cue is assigned greater weight in estimates of older speakers. When the speech is spontaneous, and hence relatively rich in age cues, the listeners seem to rely more on other cues than speech rate when estimating the age of younger speakers, whilst speech rate is still an important cue in the more difficult situation of age estimates of older speakers’ age. Moreover, the cross-experiment analysis also supported the assumption that spontaneous speech contains more age information compared to read speech, as a main effect of speech material demonstrated higher accuracy in estimates based on spontaneous speech (Figure 3).
Figure 3. Age estimation accuracy in Study 1, Experiment 1 (read speech) and Experiment 2 (spontaneous speech). The estimates are made for young, middle aged, and older speakers, based on read and spontaneous speech at natural (un-manipulated) speech rate.

Study 2

Background and aim

Witnesses of crime often provide assessments about the age of unknown perpetrators and such information can indeed be valuable in crime investigations. It is therefore important for law enforcers to have knowledge about the grounds on which age estimations are made (such as the relation between specific voice parameters and age estimates) and how precise estimations can be expected to be. In some forensic cases, interception may be performed to provide voice recordings that can be used to identify criminals through forensic voice analysis. In other cases identification may be achieved by ear witnesses. In either case, voice identification is subject to error at a relatively high rate (Boë, 2000) and may often be further afflicted by the fact that criminals frequently disguise their voices in order to obstruct identification (Boë, 2000; Neuhauser, 2008; Orchard & Yarmey, 1995; Reich & Duke, 1979; Sunetha, 2013). Voice disguise can be performed in various ways,
some of them with the help of electronic devices, others by using mechanical techniques (Perrot & Chollet, 2012). Kunzel (2000) notes that 15–25% of the cases processed at the speaker identification section at BKA (the German Federal Police Office) contained common non-instrumental forms of vocal disguise including whisper, imitation of dialect or foreign accent and age disguise with the intention to sound younger or older. In online grooming-cases, perpetrators often pretend to be younger than they are in order to attract children (Ashcroft, Kaati & Meyer, 2015). Vocal age disguise may be used by the perpetrator in cases where telephone contact is initiated. Hence, knowledge about age disguise is important to prevent the growing amount of cyber related crime, particularly in cases of online grooming. Identification of voice parameters that are resistant to disguise would have an applied value.

The purpose of Study 2 was to investigate the effect of age disguise on speech properties that are known to be related to perceived age, as well as to investigate how successful the age disguise was. That is to say, if speakers were estimated as older when they tried to sound older, and estimated as younger when they tried to sound younger. We expected to find that young speakers are able to manipulate their voices to sound older. However, we believed that middle aged and older speakers would be less successful than young speakers to disguise their voices to sound younger or older. Also, F0 is in another range for women than for men, and we therefore asked whether women and men were equally good at modifying their voices to sound a different age. Finally, it was asked if disguising the voice to sound younger was as effective as disguising the voice to sound older.

Procedures
Speech samples were produced by 36 speakers from three age groups: 20–25 years, 40–45 years, and 60–65 years. There were 12 speakers in each age group (six women and six men). The speakers in the two older age groups were instructed to sound about 20 years younger in one condition, to use their natural voice in another condition and to sound about 20 years older in a third condition. We did not include speech samples from speakers in the youngest age group disguised to sound younger because the voice condition required the speakers to try to sound like children of 0–5 years of age, which is quite another task than what was required from speakers in the older age groups and in the other voice conditions. Thus, speakers from the youngest age group were only heard in the natural voice condition and in the condition where they were disguised to sound older. Thus, the youngest age group (20–25) was instructed to sound about 20 years older in one voice condition and to use
their natural voice in another. In all, 96 speech samples from the 36 speakers were used. The acoustic analyses on speech rate and F0 were made in Praat, a software tool for analysing, synthesising and manipulating speech.

Sixty students (47 females and 13 males) with Swedish as their native language participated in the listening-experiments. The speech samples were presented to the listeners in a laboratory through headphones. The listeners were instructed to estimate the age (in years) of each speaker they heard and to write their estimates in a protocol. The listeners were randomised into three listener groups. Each group listened to 12 neutral speech samples that were produced by two female and two male speakers from each age group, 12 speech samples disguised to sound older that were produced by two female and two males speakers from each age group, and eight speech samples disguised to sound younger that were, produced by two female and two male speakers from each of the two older age groups. Hence, each participant listened to and estimated the age of 32 voices in total. The speech material differed between the listener groups with respect to in which voice condition a voice was presented. Each listener heard each voice in only one voice condition. The speech samples were presented in a randomised order within each listener group with a 10 second pause after each voice.

**Design and analyses of the voice data**

The data was computed and analyzed in SPSS 22.0 using mixed analysis of variance (ANOVA) models. Post hoc analyses were computed using the Bonferroni correction. Because the study design did not include young speakers seeking to sound younger, two analyses were performed on fundamental frequency and speech rate respectively. The first design included three voice conditions (20–25, 40–45, 60–65 years), two speaker age groups (40–45, 60–65 years) and listener gender (female, male) as within subject variables, and listener gender (female, male) as a between-subjects variable. The second analysis consisted of two voice conditions (natural, older), three speaker age groups (20–25, 40–45, 60–65 years) and speaker gender (female, male) as within subject variables, and listener gender (female, male) as a between subject variable. Mean estimated age was used as a dependent measure to calculate differences in estimated age depending on speech rate.
Design and analyses of the age perception data

The statistical analyses were conducted in SPSS 22.0 using repeated analysis of variance (ANOVA). Post hoc analyses were computed using the Bonferroni correction. Two ANOVAs were performed. The first was a within-subject ANOVA that included three voice conditions (young, natural, old) and two speaker age groups (40–45, 60–65 years). The second one was a within-subject ANOVA consisting of two voice conditions (natural, old) and three speaker age groups (20–25, 40–45, 60–65 years). Speaker gender was included as a third within-subjects variable and listeners gender as a between subjects variable in both analyses. Linear regression analyses were computed to investigate how much of the variance in estimated age F0 and speech rate explained in each of the three voice conditions. Gender was also included in the model because of its strong relatedness to F0.

Results and conclusions

The results demonstrated that the speakers increased F0 and speech rate when trying to sound younger while they decreased F0 and speech rate when trying to sound older. This strategy was applied regardless of speaker gender or age. The strategy was effective to some extent, as voices in the two disguised voice conditions obtained age estimates in the attempted direction (Figure 4). This finding held for both female and male voices, and there was no difference in effectiveness between voice disguise to sound younger and voice disguise to sound older. An interaction was found between vocal disguise and age group, such that speakers who were 60–65 years old were more successful in sounding older compared to speakers from the other age groups. However, this interaction is probably of little practical importance in that few 60-year olds would gain much from appearing older, and in absence of other interactions, we conclude that the effect of voice disguise is robust, though the effect on age estimation is rather small, typically varying from 2 to 4 years. Although speakers made linear changes in both F0 and speech rate when trying to sound younger and older, it was speech rate that explained the variance in estimated age (around 20%).
Figure 4. Effects of voice disguise, to sound younger or older than normal, on perceived age, among young, middle aged and older speakers. Error bars indicate the standard error of the mean (SEM).

Study 3

Background and aim

Accuracy in estimation of speaker age is influenced by aspects such as the chronological age of the speaker and the speech material that the estimations are based on. Such general influences can be valuable in evaluations of witness descriptions containing age assessments. In the context of witness testimony, it is also of interest to find out if listeners themselves can accurately assess the validity in their age estimations. When it comes to eyewitness identification (e.g., Sporer, Penrod, Read, & Cutler, 1995), as well as earwitness identification (e.g. Zetterholm et. al., 2012) many studies have found that the relationship between confidence and accuracy is weak. In other words, an identification made by a witness who is confident that his or her identification is correct should not be expected to be more accurate than an identification made by a less confident witness. Exceptions have been found when visual conditions are optimal (Lindsay, Read, & Sharma, 1998), although this is rarely the case at a real crime scene. Still, investigators and
courts frequently ask witnesses to rate how confident they are that their testimonies are correct on a scale from 0–100%. The realism in confidence ratings regarding identifications of faces (Brewer, Keast, & Rishworth, 2002; Juslin, Olsson, & Winman, 1996; Palmer, Brewer, Weber, & Nagesh, 2013) is well studied. However, from a witness psychological point of view, it is relevant to also examine how aspects such as speaker gender and age might affect estimation confidence and whether an individual’s confidence can be used as an indicator of that individual’s level of accuracy when it comes to estimations of personal characteristics such as age, based on vocal information. It is also of interest to explore whether conditions such as speech material (read vs. spontaneous) play a role in the degree of accuracy and confidence.

Study 3 included two experiments that investigate accuracy and confidence in estimates of speaker age. In Experiment 1, the estimations were based on spontaneous speech (the speakers gave spontaneous descriptions of how to get from a point A to a point B on a map). In Experiment 2, the estimations were based on read speech (all of the speakers read the same text). The purpose of the study was to explore differences in accuracy and confidence depending on speaker gender and age, listener gender, and speech material. Another purpose was to examine whether listeners who made more accurate age estimates, were also more confident that their estimations were correct.

**Procedures**

Speech recordings from 36 native speakers of Swedish were used in the study. Twelve of the speakers were 20–25 years old, 12 were 40–45, and 12 were 60–65. Six speakers in each age group were female and six were male. In Experiment 1, the speakers gave spontaneous walking directions on how to get from a point A to a point B on a map. In Experiment 2, the speakers read a 35-word text containing written walking directions instead of giving spontaneous directions from a map as in Experiment 1.

Sixty-five listeners (35 women and 30 men) participated in Experiment 1, and 43 listeners (24 women and 19 men) participated in Experiment 2. In both experiments, the speech samples were presented through headphones, and the listeners were asked to estimate the age of each speaker and, for each estimate, to rate how confident they were that the estimate was correct on a scale from 0–100%.
Design and analyses
In both experiments, statistical analyses were conducted in SPSS 24.0 using repeated-measures analysis of variance (ANOVA), Pearson product-moment correlations, and t tests. A 2 (speaker gender: female, male) x 3 (speaker age group: 20–25, 40–45, 60–65 years) x 2 (listeners gender: female, male) mixed design was used to measure differences in accuracy depending on speaker gender and age (within-participant variables) as well as listener gender (between-participant variable). The same analysis and design was repeated with confidence instead of accuracy as a dependent variable. Post hoc analyses were computed using the Bonferroni correction. Paired-sample t tests were computed to measure differences in accuracy and confidence depending on speech material. To measure whether individuals who made more accurate age estimations also were more confident in their estimations, correlations were computed between the mean deviation from chronological age and the mean confidence for each listener.

Results and conclusions
The results revealed that both accuracy in estimation of speaker age, and listeners’ confidence that their estimates were correct, depend on gender and age of the speaker. Accuracy was higher in estimates of male speakers’ age than in estimates of female speakers’ age for both spontaneous and read speech. In the case of spontaneous speech, higher accuracy in estimates of male speakers’ age was found across age groups (Figure 5), while it was restricted to the two oldest age groups when estimates were based on read speech (Figure 6). Although accuracy was higher in estimates of male speakers’ age, listeners were found to be more confident in estimates of female speakers’ age when speech was spontaneous (Figure 7), as well as read (Figure 8). Interactions between speaker gender and age revealed that the higher confidence in estimates of female speakers was confined to the youngest age group. It was also found that older speakers’ age was estimated with lower accuracy compared with the two younger age groups, while confidence was lowest in estimates of the middle-aged group in both experiments. Regarding listener gender, no difference was found in accuracy or confidence.
Figure 5. Accuracy (unsigned mean deviation from chronological age) in estimates of speaker age based on spontaneous speech from female and male speakers in three age groups (20-25, 40-45 and 60-65 years).

Figure 6. Accuracy (unsigned mean deviation from chronological age) in estimates of speaker age based on read speech from female and male speakers in three age groups (20-25, 40-45 and 60-65 years).
Figure 7. Mean confidence percentages in age estimates based on spontaneous speech from female and male speakers in three age groups (20-25, 40-45, 60-65 years).

Figure 8. Mean confidence percentages in age estimates based on read speech from female and male speakers in three age groups (20-25, 40-45 and 60-65 years).

In correspondence with many studies on the relationship between confidence and accuracy in eyewitness identification, confidence was found not to be an indicator of accuracy in estimation of speaker age, as correlations were close to zero and negative in both experiments. However, the results from previous studies, that estimates based on spontaneous speech are more accurate than those based on read speech, was not replicated in this study.
Discussion

The aim of this thesis was to investigate several factors related to accuracy in estimation of speaker age. The aim was also to study the role of certain speech properties in perception and manipulation of speaker age, as well as their interaction with speech material type. This was done in three experimental studies investigating different aspects of age estimation.

Discussion of main results

The relative importance of speech rate

In Study 1, one purpose was to investigate how changes in speech rate influenced listeners’ age estimates of young adult (20-25 years), middle aged (40-45 years) and older (60-65 years) speakers. Speech rate has been described as a particularly important cue to speaker age (Harnsberger et al, 2008; Schötz, 2004; Schötz, 2006; Shipp, Qi, Huntley, & Hollien, 1992; Stölten & Engstrand, 2003; Winkler, 2007) and previous studies have suggested that increased speech rate makes speakers sound younger, while decreased speech rate makes speakers sound older (Harnsberger et al., 2008; Stölten & Engstrand, 2003; Winkler, 2007). Thus, we hypothesized that decreased speech rate would increase the perceived age while increased speech rate would decrease the perceived age. In Experiment 1, where read speech was used, the results confirmed that listeners perceived speakers as younger when speech rate was increased and as older when speech rate was decreased, replicating the results of Harnsberger et al (2008), Stölten and Engstrand (2003) and Winkler (2007). Yet, it should be kept in mind that the age of younger speakers is systematically overestimated while the age of older speakers is systematically underestimated (Braun, 1996; Braun & Cerrato, 1999; Brückl & Sendlmeier, 2003; Hollien and Tollhurst, 1978; Huntley et al., 1987). Thus, decreased speech rate did not generally make young speakers sound older than their chronological age, and increased speech rate did not generally make older speaker sound younger than they actually were, as the systematical over- and underestimations of age were greater than the magnitude of the effect of speech rate. Having said that, the results from Study 1 support the assumption that speech rate is an important cue to speaker age, as a minor change by 10% in speech rate, had a significant effect on estimated age. Our results demonstrate that listeners rely on speech rate as an important age cue, and suggests that even subtle changes in speech rate has impact on perceived age. A predictable and linear decrease in speech rate has been found for both
women and men in studies on age related changes in speech (Bürckl & Sendlmeier, 2003; Linville, 2001; Oyer & Deal, 1985; Smith, Wasowicz, & Preston, 1987). The general decrease in speech rate with ageing can probably explain the reliance on speech rate in age perception to some extent. Of course, we do not claim to have established that speech rate is more important than other specific cues, as we did not measure the effect of speech rate relative to other speech properties. Yet, the fact that the minor change in speech rate had an effect on perceived age, in absence of changes in other speech properties, confirms that it is one of the properties that listeners extract and infer speaker age from, especially when age estimates are based on read speech. Another purpose of Study 1 was to investigate whether the impact of speech rate depended on speech material. That is, whether speech rate is less important when age perception is based on spontaneous speech than on read speech. Because read speech does not contain linguistic variation, listeners must rely more on non-linguistic age cues, such as speech rate, than they would have done if they had access to linguistic cues. We wanted to investigate whether linguistic variation added age information, and thus, made the listeners’ age estimates more accurate. We found that the listeners relied less on speech rate as a cue to speaker age when spontaneous speech was used, than when read speech was used. This suggests that listeners use non-linguistic cues to speaker age to a lesser extent, when the speech material is richer in linguistic age cues. The result might even imply that listeners prefer linguistic cues. Yet, spontaneous speech and read speech differ in many ways, not only concerning linguistic information. Spontaneous and read speech also differ in several acoustic cues such as prosodic and spectral information (Howell & Kadi-Hanifi, 1991; Nakamura, Iwano, Furui, 2008). Furthermore, the boundaries between tone units differ between spontaneous speech and read speech (Blauw, 1994), the position of the stresses are different and there are also fewer pauses in read speech than in spontaneous speech (Howell & Kadi-Hanifi, 1991). Moreover, spontaneous speech has a more constrained spectral space (Nakamura et al., 2008). These factors may further explain why the interaction between speech rate and chronological age, in the present study, was slightly different depending on speech material type. Furthermore, it was found that in estimates of older speakers’ age, listeners used speech rate as an age cue even when the speech was spontaneous. One reason might be that the age of older speakers is harder to assess, as confirmed by the accuracy results in Experiment 1 (in Study 1). Another reason could be that age related changes in speech rate are actually particularly pronounced in older age, due to physiological and cognitive changes (Burke & Shafto, 2004; Bürckl &
Sendlmeier, 2003; Linville, 2001; Oyer & Deal, 1985; Schötz, 2006; Smith, Wasowicz, & Preston, 1987). When read speech was used, speech rate had an effect on perceived age regardless of chronological age of the speakers, but the effect was smaller in estimates of younger speakers compared to that of older ones. Thus, it was concluded that the importance of speech rate seems to increase with the chronological age of the speaker. To summarise, we tested if different age estimation strategies are used depending on speech material and could confirm that it was the case, regarding speech rate, but that age estimation strategy was also dependent on speaker age.

Vocal and perceptual effects of age disguise

In Study 2 one aim was to investigate the effect of age disguise on speech rate and F0, which are speech properties known to be related to perceived age (Linville, 2001; Shipp, qi, Huntley & Hollien, 1992; Stölten & Engstrand, 2003). Another aim was to investigate how successful the age disguise was. That is to say, if speakers were estimated as older when they tried to sound older, and estimated as younger when they tried to sound younger. The results revealed that the speakers increased both F0 and speech rate when they were trying to sound younger, while they decreased both F0 and speech rate when trying to sound older. Although actual changes in age related speech properties are dependent on age and gender, the speakers changed F0 and speech rate in the same way, regardless of gender and age. Previous studies have illustrated that listeners’ estimates of speaker age are affected by age stereotypes (Hummert, Mazloff, & Henry, 1999; Hummert, Shaner & Gartska, 1998). It is likely that that the speakers in Study 2 based their speech manipulations on both more or less stereotypical ideas about ageing and on actual experiences of how female and male voices change with age. However, the age disguise turned out to be somewhat successful, as the listeners perceived the speakers as older when the speakers were trying to sound older, and vice versa. Thus, the speakers were able to manipulate speech properties related to perceived age to some extent, which replicates findings from Hautamäki, Kanervisto, Hautamäki and Kinnunen (2018), as well as Lass et al. (1982). This finding held for both female and male voices, and there was no difference in effectiveness between voice disguise to sound younger and voice disguise to sound older. However, there was an interaction between voice disguise and age group, such that speakers 60–65 years old were more successful in sounding older than speakers from the other age groups. These results reflects the fact that different age cues are important in different age groups, and that some vocal features are not as easily modulated as others. In
In this study the acoustic measurements were restricted to F0 and speech rate, while speakers are likely to modulate several speech properties to disguise age. It was concluded that speakers changed F0 and speech rate in order to disguise their voices to sound younger and older, and that age disguise caused a change in perceived age in the intended direction. However, in accordance with Lass et al. (1982), the effect on age estimation was found to be rather small, generally varying from 2 to 4 years. One reason for the rather small effects might be that age related changes in the speech production mechanism sets boundaries for how much speakers are able to change their voices in order to disguise age. As it was found that the speakers did not change F0 in a manner that corresponds to actual age related changes in F0, we might also suggest that the small effect of age disguise might be partly explained by the fact that the speakers’ strategies to modulate perceived age did not fully resemble actual age related vocal changes, or at least not those related to perceived age. The acoustic analyses showed that speakers made linear changes in both F0 and speech rate when trying to sound younger and older. Yet, it was speech rate that explained the variance in estimated age (around 20%). The suggestion by Bürckl & Sendlmeier (2003) that speech rate is a more important perceptual cue to speaker age than F0 is, was confirmed.

**Accuracy and confidence in estimation of speaker age**

In Study 3 the aim was to investigate differences in accuracy and confidence depending on speaker gender and age, listener gender and speech material. Another aim was to investigate the relationship between accuracy and confidence. That is, whether the listeners’ confidence that their age estimations were correct, indicated how accurate the estimations actually were. It was concluded that confidence should not be used as an indicator of accuracy as the correlations between confidence and accuracy were negative and close to zero, regardless of speech material. Weak relationships between confidence and accuracy have been reported in several studies on identification of faces (Brewer, Keast, & Rishworth, 2002; Juslin, Olsson, & Winman, 1996; Palmer, Brewer, Weber, & Nagesh, 2013) and one could argue that a low confidence-accuracy relationship should be expected in age estimation as well. Yet, estimation of speaker age is a quite different task than identification in important ways. While identification demands memory and recognition of a previously observed individual, age estimation is often performed simultaneously as, or directly after, the presentation of the target voice and demands use of perceptual cues and social categories.
It is likely that age estimation to a large extent, depends on implicit perceptual processes. Thus, listeners might not necessarily have been aware of what cues they based their age estimates on, and whether some age estimates were more reliable than others were. Some speakers that were less accurate might have used a more explicit age estimation strategy, which was not adaptable for all speakers, as individual variation in speech properties between different speakers is large, but that made them feel more confident that their estimates were correct. Correspondingly, some listeners who relied more on implicit perceptual cues without using an explicit age estimation strategy might have assessed the grounds for their estimates as less reliable. A related finding was that the gender patterns of accuracy were different than the gender patterns of confidence as listeners were more confident but less accurate in their estimates of female speakers’ age, but less confident and more accurate in their estimations of male speakers’ age. Reversed gender patterns for confidence versus accuracy was found for both read and spontaneous speech. These seemingly contradictory results might be related to the fact that the female and male voices age in different ways (Dehqan, Scherer, Dashti, Ansari-Moghaddam, & Fanaie, 2012; Hirano, Kurita & Sakaguchi, 1989; Linville 2002). Again, if listeners adopted an explicit age estimation strategy, that might have been more adapted for male speakers’ age, but applied it on estimates of both female and male speakers, it is reasonable that accuracy is higher in estimates of male speakers. Yet, these are only speculations, and it does not explain the higher confidence in estimates of female speakers’ age. It should also be noted that the role of speaker gender in age estimation accuracy is unclear. The higher accuracy in estimates of male speakers’ age found in Study 3 contradicts the findings of Harnsberger et al (2008) and Neiman and Applegate (1990), who found that listeners were more accurate in estimates of female speakers’ age. However, the speculations made here should be further explored in a study designed to investigate age estimation strategies related to speaker gender and whether accuracy benefits from implicit or explicit processes. Of course, confidence might have been influenced by other aspects, such as personality traits, unrelated to age estimation skills. To make things even more complicated, the results in Study 3 also demonstrated that both accuracy and confidence were dependent on interactions between age and gender of the speaker. In the case of spontaneous speech, the higher accuracy in estimates of male speakers’ age was found across age groups, while it was restricted to the two oldest age groups when estimates were based on read speech. Interactions between speaker gender and age revealed that the higher confidence in estimates of
female speakers was confined to the youngest age groups when spontaneous speech was used. Corresponding with previous studies (Huntley et al., 1987; Jacques & Rastater, 1990; Moyse et al., 2014) it was also found that older speakers’ age was estimated with lower accuracy compared with the two younger age groups. Yet, confidence was lowest in estimates of the middle-aged group in both experiments. Regarding listener gender, no difference was found in accuracy or confidence. It was concluded that although accuracy was higher in estimates of male speakers’ age, listeners were found to be more confident in their estimates of female speakers. It was also concluded that the relationship between accuracy and confidence was very weak (close to zero in both experiments). However, the indications in Bürckl & Sendlmeier (2003) and in Study 1, that estimates based on spontaneous speech are more accurate than those based on read speech, could not be replicated in Study 3. One obvious question raised by some of the more unexpected results in this study is how the confidence rating, following each age estimation, might have affected the age estimation process.

The results of study 3 raise many questions, but also delivers a clear answer to the question about whether confidence can be used as an indicator of accuracy. For both spontaneous and read speech, it was concluded that the answer was no.

**Methodological issues and future studies**

Based on the results in Study 1, it was concluded that speech rate is an important cue to speaker age, particularly when age estimates are based on read speech. However, we did not study the importance of speech rate relative to other speech properties, and therefore we cannot draw conclusions about how important speech rate is compared to other speech properties, nor about how it correlates with other speech properties. However, one purpose of the study was to investigate whether the assumption that speech rate is particularly important, also applies when speech is spontaneous. Of course, the degree of spontaneity in our spontaneous speech samples could be questioned as all speakers were recorded while giving walking directions from a map. Thus, the speakers could not choose freely what to talk about. Yet, we found it important that all speakers talked about the same subject, as we wanted to avoid that the topic in the speech interfered with the age-related information we wanted to study. We also considered that some speakers might have experienced a completely spontaneous speech task as stressful. However, it would have been beneficial to use a third type of speech material with a higher degree of spontaneity, for example conversational speech, to
make it possible to investigate how degree of spontaneity is related to accuracy and to the importance of different speech properties. Furthermore, it has been demonstrated that voice identification is facilitated by longer speech samples rather than by greater vowel variability (Cook & Wilding, 1997). It would be meaningful to study the role of speech material length in estimation of speaker age, using a similar study design.

In study 2 we found that speakers modulated both F0 and speech rate in order to sound younger and older and that the disguise had an effect on perceived age in the intended direction of the disguise. These results have implications for forensic detection of voice disguise. However, to get a more complete picture of how vocal age disguise is performed, several speech properties, such as spectral features, harmonics to noise-ratio and VOT should be measured.

Turning to Study 3 we found that the correlation between confidence and accuracy was weak and possible perceptual and vocal explanations were discussed. However, we should also consider methodological explanations that might have contributed to the weak correlation. It might be the case that the age estimation task was not appropriate when it was followed by a confidence rating. In spite of the downsides of using age categories in age estimation tasks, it might have been more adaptive in Study 3, where a confidence rating was related to each age assessment. We asked the listeners to make age estimations in exact years and to rate their confidence in each estimate on a scale from 0-100%. Of course, the chances of making a perfectly correct age estimate were small, and different listeners might have understood the confidence task in different ways. Some listeners might have interpreted “correct” as the exact chronological age of the speaker in years, while others might have interpreted “correct” as close to the exact age. By using age categories instead, this issue could have been avoided, as the listeners would then have had a reasonable chance of making perfectly correct age estimates.

Concluding remarks
The studies in this thesis have contributed to knowledge about some aspects of estimation of speaker age. Although the studies do not answer overarching questions such as what voice properties are the most important ones, it provides some insight into factors that influence perception of speaker age on which further research can unfold. We have shown that even subtle changes in speech rate have an effect on age perception. One important finding in this thesis is that the effect of speech rate depends on the speech material used;
read speech or spontaneous speech. Our results suggest that when linguistic cues are absent, listeners rely more on speech rate than they do when the speech is spontaneous and they are able to also use linguistic cues to speaker age. Thus, the study implies that the speech properties that are most important in age perception in everyday life, might not be the same ones as those found to be the most important ones when read speech is used in highly controlled listening experiments. We have also shown that speakers are able to modulate their speech to sound younger and older, even if the effects of the age disguise on perceived age was small, and dependent on speaker age. The speakers spontaneously manipulated speech rate and fundamental frequency to sound younger and older. Yet, the listeners relied more on speech rate than on F0 in their age estimates. These results may contribute to forensic strategies for detection of age disguise in recorded material, even if much more research would be needed to provide a complete picture of vocal age disguise. Another finding made in this thesis, which is of forensic relevance, is that estimator confidence shall not be used as an indicator of estimation accuracy. That is, listeners who are more confident that their estimates of speaker characteristics, such as age, are correct are not necessarily more accurate than those who are less confident.

The findings we have made is just a small part of a much larger puzzle, but the insights made in this thesis can hopefully be of use in further investigation relating to age perception.
References


