LUND
UNIVERSITY

Department of Psychology

When is hearing believing: The effects of professional experience on voice identification accuracy

Benjamin Scimone

Master of Science Thesis SS 2013

Supervisor: Mats Dahl

Abstract

The purpose of the study was to examine differences in the accuracy of voice identification tasks between an experimental group (N = 56) consisting of security personal, and a control group (N = 44). In addition, calibration methodology was employed in order to examine if the experimental group exhibited a higher confidence level in their identifications than the control group. No significant differences between the two groups were found on voice identification tasks, nor were any significant differences in realism of confidence between the two groups established.

Key words: Earwitnesses, Voice identification, Face Identification, Calibration, Confidence, Witness-psychology, Formal training, Informal training.

Table of Contents

Introduction

Researchers within the field of earwitness psychology has advocated for more research into the earwitness process due a tilted scientific focus towards eyewitness psychology, especially in light of the translucency with which DNA exposes the pit-falls of judicial systems contingent on witness testimony. The impact of technology such as DNA, and case studies conducted in the United States have confirmed that eyewitness testimony have contributed to more wrongful convictions than all other causes combined (Huff, Rattner & Sagarin, 1986). Considering the numerous crimes committed by masked perpetrators, and/or under conditions when visibility is limited, or the spoken threats, and harassments perpetrated over the telephone, it becomes apparent that factors unique to the earwitness process are important scientific objectives (Clifford, 1980, Yarmey, 1995). At present, earwitness-psychology is but a small, and budding portion of the broader field of eyewitness-psychology, however, considering the severe, and long-term consequences associated with wrongful convictions, and the cost that criminality poses to society at large, earwitness research is a scientific endeavor of increasing relevance.

Earwitness testimony

Jacob in the Bible is likely the first written reference to an earwitness (see the book of *Genesis*), while William Hulet who in England 1660 was accused of the murder of a king (Charles I), is the first known person to be convicted by a court on the basis of earwitness testimony (Hollien, Bennett & Gelfer, 1983). Empirical research of the psychological processes governing earwitness testimony has a shorter pedigree, and can be traced to a

prominent trial in the 1930s. The celebrated American aviator Charles Lindbergh's infant son was abducted from the family home in March 1932. Lindbergh is subsequently contacted by a person claiming to have kidnapped his child, following a transaction Lindbergh is in a position to overhear the abductors voice. The negotiations, however, does not end in the child recovery, instead the infant's body is found a few months later. A German immigrant named Bruno Hauptmann is arrested, and Lindbergh is called to testify for the state in 1935. His testimony essentially conveys that Hauptmann's voice was the one he'd heard in the spring of 1932. It is generally conceded, that the earwitness testimony of Lindbergh was decisive in making the prosecution's case. Hauptmann was convicted of kidnapping, and murder, and sentenced to death. He was executed in New Jersey in, 1936. The Lindbergh trial attracted a substantial amount of media attention for the time, and questions regarding the veracity of earwitness testimony arose. Although controversial, the trial also served to ignite scientific interest into the capacity of earwitnesses, and any subsequent credence which should be attached to earwitness testimony in general by the judicial system (Yarmey, 1995, Broeders, 2001, Solan & Tiersma, 2003). In 1937, McGehee conducted the first in a series of experiments examining peoples aptitude for identifying an unknown speaker. Her first experiment, showed that after 5 months, the identification capacity of her subjects were down to a mere 13 %, in 1944, she performed a replication with refinement of her first experiment, adding features that were more applicable to actual legal situations involving voice identification. McGehee's findings in 1944, were similar to her first experiment in 1937 (McGehee, 1937, 1944). Never-the-less, since McGehee's pioneering work, only a handful of studies has been conducted within the earwitness psychological field.

In legal settings there are generally two forms of witness confrontations used in eyewitness testimony, the simultaneous, and the sequential, the former is the most commonly used, and entails that the suspect, and foils (persons/voice samples that are present merely for

distractional purposes) are presented to the witness simultaneously. The sequential as implied by the name means that the suspect, and the foils are presented one after the other or in sequence (Broeders, 2001). The most common situation applied to earwitness line-ups will be an earwitness that on one prior occasion has heard an unknown voice, and is exposed to a sequential voice-parade (Granhag, 2001). Usually witnesses are eyewitnesses, relating information pertaining to something they've experienced visually (Broeders, 2001), and procedures used for earwitness line-ups are firmly anchored in those used for eyewitnesses. However, the preconditions for making an identification based on an auditive compared to a visual stimulus is very different.

Recognition of speech

The distinctive nature of our speech is apart from the biological determinants, a product of our accents, rhythms, and adopted behaviors, such as pauses, inflections, and the pronunciations of certain words, and phrases. The use of speech as a means when identifying a person is dependent on what is referred to as *interspeaker* variability, which is assumed to be greater than the *intraspeaker* variability, when hearing the same voice on different occasions (Hammersley & Read, 1989, Yarmey, 1995). This essentially means that we observe smaller differences between several speech samples given by one person (intra-speaker) on several occasions than we'd experience when comparing two samples originating from two different individuals (inter-speaker). The interspeaker variability may be affected by speech disorders such as stammering, and stutters while the intraspeaker variability is subject to change depending on context demands, intentions, mood, emotion, and distractions of the speaker (Yarmey, 1995). It's unlikely that a speaker will pronounce a single word exactly the same way, even if repeating it in succession (Hollien, 1990).

When listeners engage in voice identification a range of characteristics serve to facilitate their classification of that voice. Subsets of auditory cues such as, levels, pattern, and variability of pitch, vocal intensity patterns, dialect, articulation, voice quality, and prosody (rhythm, and or melody of speech) (Hollien, 2002). It is believed that if one characteristic fails in being useful for the identification/recognition process another parameter may serve if it's distinct enough (Lavner, Gath & Rosenhouse, 2000). It has been suggested that voices differ from each other in other ways, for example; in (a) voice quality, (b) voice dynamics (e.g. prosodic features of speech, rhythm, rate, continuity, and intonation), (c) pronunciation, (d) vocabulary, and (e) style (Sapir, 1927 but see Yarmey, 1995). One proposal of what listeners attempt to focus on in the recognition process is described as follows; (a) fundamental speech frequency or perceived pitch, (b) articulation, (c) general voice quality, (d) prosody, (e) vocal intensity, (f), and certain characteristics of speech such as, dialect, linguistic stress, peculiar speech patterns, and speech impediments (Hollien, 1990). For instance, Thompson (1987) tested monolingual English speakers on voice identification using speech samples provided by bilingual speakers. After a 1-week retention, Thompson found that the test subjects were far better at identifying those speakers who spoke English fluently, followed by speakers speaking English with a heavy Spanish accent, and the worst identification rates were produced when the target voice was speaking Spanish. Thompsons finding has since been replicated in at least four experiments (Goggin, Thompson, Strube, & Simental, 1991). There is generally consensus regarding the positive effects of language familiarity on voice identification (Hollien, Majewski & Doherty, 1982). It has been suggested that the reason people have more difficulty identifying or discriminating between accented voices may be a function originating from the listener rather than the speaker. Just as the findings in eyewitness testimony which have established that it's easier to discriminate, and identify people of the same ethnicity. A similar process may be at play when we hear a

voice unfamiliar to our ethnicity, region or dialect (Ladefoged & Ladefoged, 1980). In earwitness research a distinction is made between merely recognizing a voice, and identifying it. Speaker or voice recognition refers to the activity of examining if an earwitness has familiarity with a voice sample (Brown, 1979), while speaker identification is related to an earwitness attempting to determine if a given voice, usually in an array of several voice samples, is identifiable as the one originally heard on a prior occasion, for example in connection with a criminal event experienced by the witness (Yarmey, 1995). In the current study the voice samples used are provided by speakers unknown to the participants. There are some very fundamental differences associated with the recognition process that people employ for familiar voices compared to unfamiliar voices. Several factors, such as greeting phrases, favorite expressions, nicknames, stereotypical phrasing, and peripheral context-related information ensures that its often an unique pattern of perceptual factors at work when we're hearing a voice known to us (e.g.,Van Lancker Kreiman, & Emmorey, 1985, Van Lancker, Kreiman, & Wickens, 1985). Under optimal conditions 2 seconds or less seem to be sufficient for people to recognize a familiar target voice. It has been shown that the voices of colleagues, friends, and family, in addition to, celebrities such as politicians, radio hosts, and television personalities are recognized with high accuracy based on a limited speech sample (Bricker & Pruzansky, 1966 as cited in Yarmey, 1995). Peoples capacity for voice recognition alters dramatically if we use unfamiliar voices as target voices opposed to familiar ones, Hollien (2002) suggests that in order for a voice to be considered familiar, the listener needs unimpaired hearing, and listened to the voice more or less regularly for a period of no less than 2 years.

It has been well established that the memory for an unfamiliar voice tends to be stronger briefly following exposure to it, and the recognition of it generally declines over retention periods up to 5 months (Hammersley & Read, 1985, McGehee, 1937, Clifford, Rathbone &

Bull, 1981). Several factors influence any effects that delay may have on voice memory, such as planned versus unplanned, selective attention during the listening phase (Saslove & Yarmey, 1980), and the uniqueness of the voice sample in comparison to subsequent voices (Yarmey, 1991).

Another important factor is age. The seemingly superior age for voice recognition has been observed to be between 21–40, where people generally outperform older subjects (Bull & Clifford, 1984). Hearing loss becomes more prevalent as people age, especially for high frequency sounds it's likely that voice identification is negatively correlated with age (Yarmey, 2000), other forms of hearing impairments are also generally associated with advancing age, and may explain why people tend to do worse at aural perceptual tasks beyond the age of 40 (Popelka, Cruickshanks, Wiley, Tweed, Klein & Klein, 1998).

Formal and informal training of voice identification

The scientific findings of the effects of informal training on voice identification is very limited. A representative of applied linguistics such as phonetician, may conduct what is generally referred to as forensic speaker identification, and compare voice samples in an attempt to identify an offender or eliminate a suspect. Although phoneticians practice voice identification, and can, when supported by formal training, identify many characteristics of a voice, forensic speaker identification remain a highly convoluted process where no known set of criteria's can be used reliably to identify a speaker (Nolan, McDougall, de Jong & Hudson 2006), as mentioned earlier the intraspeaker variability can be extensive, aggravating the identification process considerably. The scientific findings of the effects of informal training on voice identification is relatively limited. There are some indications that blind earwitness

are slightly more proficient at voice identification than seeing counterparts (Bull, Rathborn & Clifford, 1983), however, it has been proposed that being a good listener or having good encoding stratagems facilitate voice identification to a greater degree than training by itself (Clifford, 1980).

Confidence and Calibration

The perceived veracity of a witnesses statements often rests on how confident the witness appears to be (e.g. "I *know* what I saw/heard..." or "I am certain beyond any doubt that..."). Research has shown that witnesses tend to exhibit overconfidence in their ability for recall of a memory (Allwoord & Granhag, 1999), coupled with a limited awareness of the factors that influence their cognitive processes (Charman & Wells, 2008). In judicial settings judges, and other legal representatives base much of their perception of overall witness credibility on the level of confidence elicited by the witness (Granhag, 2001). Therefore, it's important to explicate, and examine the actual relationship between confidence, and the statements that a witness makes, which is generally referred to as the Confidence-Accuracy Relationship. In many studies where an unfamiliar speaker has been used as an identification task the correlation between confidence, and accuracy has been non-significant or low (e.g. Saslove & Yarmey, 1980). There are several conditions that may exert an influence or moderate the confidence-accuracy correlation, the duration of the voice sample, quality of the speaker's voice, and familiarity with the spoken voice. In a study by Yarmey, and Matthys (1992), a significant negative confidence-accuracy relationship was found when short (18 seconds) sample durations were used. When research is conducted concerning peoples judgments, the prevailing method used is that of calibration (e.g. Björkman, 1994; Gigerenzer, Hoffrage & Kleinbölting, 1991, Juslin, Olsson & Björkman, 1997). Calibration refers to participants

realism of confidence, in other words, the relationship between a participants subjective probabilities, and corresponding objective probabilities. When calibration is applied in witness research the participants are tasked with reporting a subjective probability that the person(s) that they've identified is the target person/voice for example, this probability is commonly a percentage adopted to a 100 point scale (e.g. 11 possible probabilities between 0–100 %, 0, 10, 20, ..., 100). In an actual experiment a query might read "Report the ...% of confidence that you experience in identifying this particular person as the one that you saw earlier." The subjective probabilities of the participants are then compared with the corresponding relative frequencies of correct identifications within the confidence categories. A test subject is considered to possess realism of confidence or be well calibrated in their confidence, if all identifications made when 100 % subjective probability is reported were correct, and if 75 % of all identifications made when 75 % subjective probability was reported were correct, and so on (Olsson, Juslin & Winman, 1998). In addition, the direction of the realism of confidence is relevant in witness-psychology, and while calibration allows for assessments of the goodness of fit between confidence, and accuracy, a participants overall directionality of confidence (being over or under confident) elucidates the nature of a participants confidence-approach to, for example, a given identification task (Allwood & Johansson, 2004).

Research hypotheses

The objective of this study was to examine the identification accuracy, and the levels of confidence, in order to observe if they were dependent on occupational experience of such tasks. The experimental group consisting of security guards, emergency operators, correctional facility staff, and police officers have years of occupational experience of varied levels of voice identification tasks, it is therefore predicted that they'll have a higher accuracy, rate on their voice identifications, and be more well calibrated on their judgments. Furthermore, facial identification tasks were added as a control condition to examine if any possible differences for eyewitness identification was noticeable between the groups.

H_{I} Occupational experience of voice identification tasks lead to a better overall identification accuracy for voices.

H_{II} Occupational experience of voice identification tasks lead to a more calibrated confidence judgment relating to voice identification tasks.

Method

Participants

The study is a quantitative study, and based on a statistical analysis of numerical data. It is important to note the study´s ecological validity, the test although devised to examine the participants capacity for voice identification can´t be presumed to equate a real life situation of voice testimony. During a genuine voice identification it would be safe to assume that a number of both internal, and external factors of both mundane, and extra ordinary character would be present, and exert influence on the would-be witness, in the experimental setting used such factors were controlled for. The study included 100 participants (N = 40 females, N = 60 males) divided into two groups with an overall age range of 19–66 years (M = 36.88). The majority of the participants (N = 97) reported Swedish as their native language, and the remaining (N = 3) reported another language as their native. In the experimental group 34 participants reported Scanian to be there spoken dialect, and 22 reported another dialect. In the control group 22 participants reported Scanian as their spoken dialect, and 22 reported another dialect. The Experimental group was recruited from a professional educational conference of security personnel. It consisted of female (N = 13), and male (N = 43) security personnel with an age range of 25–63 (M = 40.27) years, and represented an experienced group of (M = 14.46 years) professional security staff, and was generally believed to have experience of but lacking any formal training in voice identification tasks. The Control group was primarily recruited through Lund University, Sports clubs in Malmö, and Private schools in Scania. It consisted of female (N = 27), and male (N = 17) participants with an age range of 19–66 (M = 32.57), the participants in this group was generally believed to lack professional experience, and any formal training in identification tasks.

Material

The test battery consisted of 10 films, and 60 voice readings randomized into 10 experiments that had an identical design. Each experiment included 4 film sequences (lasting 10–15 sec per film) (2 female targets, and 2 male targets), 4 voice samples (lasting 10–15 sec per sample) (2 female targets, and 2 male targets), and 8 filler tasks. Furthermore, each experiment included 1 female, and 1 male film target absent, in addition to 1 female, and 1 male voice target absent condition. The filler tasks used were logical tests. The 10 films included in the experiment was captured with a tripod mounted digital camera (Brand Minolta, 3.3 megapixel 2056 x 1544). In each film sequence only 1 person was shown at a time moving in front of the camera at a distance of 2–6 meters performing a mundane activity, the target person was always present in the picture, and no distracting elements such as other people or activities were included in the film sequences. Generally the pictures used as foils (non target persons) were taken at 1,5 meters distance, and had a similar light setting to those used for the target persons. The pictures displayed in the experiments were approximately 20 x 15 cm, and had been selected to match the target persons gender, and approximate the targets age, and overall appearance.

The audio stimuli contained a voice sample reading a structured text lasting approximately 15 seconds. The readers voice was emotionally neutral. Following a filler task the test subjects heard 6 voice samples reading the same structured text in succession (simulating a sequential voice parade), with 5 second pauses between each reading. The voice readings had been recorded only once meaning that if the target was present in the condition the identical voice reading would be heard by the test subjects during the voice parade. The foils (non

target voices) were of the same gender as the target, and selected on the basis of approximation in age, and dialect to that of the target voice.

Procedure

The experiments were performed at daytime hours, in classrooms at conference facilities as well as the facilities of Lund University, with an average of 10 participants taking the test at a time. Prior to introducing the test itself the participants were informed of the general nature of the experiment. The participants were also reminded to report their level of confidence concerning the accuracy of the answers that they gave in percentages, and was informed of the parameters that these percentages would be judged upon (there would be seven possible answers to submit, six identifications, and a target absent choice per identification task. The threshold set for a reported guess was set at 14 %, 100 % / 7 ≈ 14 %). Then printed answer submission forms were handed out. Furthermore, prior to, and during the experiment the participants were informed that it was possible that the target person/voice may not be present in the identification line up.

The test started with either a film sequence or a voice sample, containing the target person or voice. After the introduction of the target stimuli a distraction task followed, and the subjects were informed that they had 60 seconds to solve the task at hand. Following the distraction task either six photos were displayed for 45 seconds or 6 voice clips were played to the participants, each lasting approximately 15 seconds, and played with 5 second intervals. The participants were only allowed to take part of the test stimuli once, at the end of the experiment the participants were reminded to fill in the last page of the printed form containing additional information about their person after which the forms were collected.

Results

The findings in this study did not support the first research hypotheses. No significant differences were found between the experimental group, and the control group in the voice identification task, .54(.25) for the experimental group, and .52(.23) for the control group. Nor do the results support the second research hypothesis, confidence judgments .65(.20) for the experimental group, and .60(.14) for the control group, and calibration .21(.14) for the experimental group, and .19(.11) for the control group. There were no significant differences for the control condition of facial recognition .38(.25) for the experimental group, and .40(.29) for the control group.

Table 1. shows frequency of identification responses for voice, and face identification tasks for the two groups. The response categories are as follows; "Hits" indicate that a correct answer has been provided when the target was present, "Misses" indicate that that a incorrect answer has been provided when the target was present, "False alarms" indicate that an identification has been made when the target was absent, "Correct rejections", indicate that the target has been identified as non-present when non-present.

Table.1 Frequency of identification response for voice, and face identification tasks.

	Hits	Misses	False alarms	Correct rejections
Voice identification				
Experiment group	.82	.18	.75	.25
Control group	.70	.30	.67	.33
Face identification				
Experiment group	. 45	.55	.70	.30
Control group	.42	.58	.69	.31

Table.2 Correct response, Confidence, Calibration, and over/under confidence.

	Correct response	Confidence	Calibration	Over/Under confidence
Voice identification				
Experiment group	.54(.25)	.65(.20)	.21(.14)	.15(.26)
Control group	.52(.23)	.60(.16)	.19(.11)	.08(.24)
Face identification				
Experiment group	.38(.25)	.65(.16)	.24(.16)	.22(.24)
Control group	.40(.29)	.56(.16)	.23(.14)	.15(.29)

Discussion

The voice samples used in this study was recorded once, with everyday speech in mind. Meaning that any stammers, miss-readings of the structured text, and the occasional peripheral sound bite was a part of the test stimuli as it was presented to the participants for identification purposes. The filler task only lasted for 60 seconds before the identification task was at hand. Notwithstanding, the voice identification performances of the participants in the current study, one can conclude that identifying an unfamiliar voice, even if you possess prior experience of similar tasks, is a difficult, and challenging endeavor for most people. Had this been an authentic earwitness situation, the voice one heard initially likely would've differed from the one presented during a voice parade, months may have passed since one first heard the voice under very different circumstances. During the process of conducting this study the author was able to have several conversations with the participants, indeed, many were intrigued, and surprised by the relative difficulty that the tasks presented. In line with the findings in prior studies there seemed to be a wide-held belief in the accuracy of voice recognition, supported by the apparent ease by which we distinguish between familiar voices. The participants in the experimental group explained that there had been a shift in the general nature of the duties prevalent in the security business over recent years. For example, it's becoming increasingly more common that security objects are manned by one person sitting in front of a monitor, responsible for allowing people access to secured areas augmented by a closed circuit television system (e.g. security camera monitor) but oftentimes on voice identification alone.

The threat of armed robbery is a very real one, when you work in the security business, and modern occurrences of such crimes readily assumes that the offenders are wearing disguises, leaving voice recognition a possible avenue for witness identification.

The present study did not find any support for the facilitation of voice identification due to occupational experience of such tasks. However, all the professions included in the experimental group have voice identification tasks as an occupational occurrence but some of the professions experience it to a higher, and lesser degree. Perhaps, the experimental group should've been streamlined to include emergency operators exclusively, who as a group has the highest expected occurrence of voice identification,. In the current state, some of the participants in the experimental group may have had limited experience of voice identification, and others a relatively great deal of experience. The aptitude for voice identification is known to decline at the age of 40, considering that the experimental group had a mean age of 40, may have influenced their performances, and countered any effect gained from professional experience. The results did not support the hypotheses that the experimental group were more calibrated in their confidence. One would expect that experience of voice identification would lead to an appreciation of the difficulty that such identifications pose, and this insight may, in turn, lead to a better calibration, or possible under confidence. The control group on the other hand was expected to display a mild overconfidence although the result did not support such a prediction.

Considering the general belief that people seem to have in the accuracy of voice identification, and the importance, and trust that criminal courts place, and operate on concerning witness testimony, it's important that research examine under what conditions, and with what certainty people can identify voices. In addition, it's relevant to investigate if experience of voice identification facilitates accuracy, and/or calibration of confidence.

The author would like to extend his heartfelt gratitude to Mats Dahl, for support, guidance, and encouragement during the course of this project, and especially for recognizing, and supplying kindling to the authors enthusiasm.

References

Allwood, C. M. & Granhag, P. A. (1999). Feelings of confidence and the realism of confidence judgements in everyday life. *Judgment and decision making Neo-Brunswikian and process tracing approaches* (pp. 123–146). Hillsdale, N.J.: Lawrence Erlbaum Press.

Allwood, C. A., & Johansson, M. (2004). Actor–Observer differences in realism in confidence and frequency judgments. *Acta Psychologica*, 117, 251–274

Björkman, M. (1994). Internal cue theory: Calibration and resolution of confidence in general knowledge. *Organizational Behavior and Human Decision Processes, 58,* 386-405.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America, 40,* 1441-1449.

Broeders, A. P. A. (2001) Interpol, Forensic Science Symposium. Forensic Speech and Audio Analysis Forensic Linguistics. Collected November 18, 2012 from www.taracentar.hr/attachments/interpol_forensic.pdf

Brown, R. (1979). Memory and decision in speaker recognition. *International Journal for Man-Machine Studies, 11,* 729-742.

Bull, R., Rathborn, H., & Clifford, B. R. (1983). The voice-recognition accuracy of blind listeners. *Perception 12,* 223 – 226.

Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. *Eyewitness testimony: Psychological perspectives* (pp. 92-123). New York: Cambridge University Press.

Charman, S. D., & Wells, G. L. (2008). Can Eyewitnesses Correct for External Influences on Their Lineup Identifications? The Actual/Counterfactual Assessment Paradigm. *Journal of Experimental Psychology: Applied*, 2008, 14(1), 5–20.

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

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

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506-528.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19,* 448—458.

Granhag, P-A. (2001). *Vittnespsykologi.* Lund: Studentlitteratur.

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

Hammersley, R., & Read, J. D. (1989). Relevance of voice identification research to criteria for evaluating reliability of an identification. *Journal of Psychology, 123,* 109-119.

Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics, 10,* 139–148.

Hollien H., Bennett, G., & Gelfer, MP. (1983). Criminal Investigation Comparison: Aural versus Visual Identification Resulting from a Simulated Crime. *Journal of Forensic Sciences,* 28, 208–222.

Hollien, H. (1990). *The acoustics of crime.* New York: Plenum Press.

Hollien, H. (2002). *Forensic Voice Identification*. San Diego: Academic Press.

Huff, R., Rattner, A., & Sagarin, E. (1986). Guilty until proven innocent. *Crime and Delinquency, 32,* 518–544.

Juslin, P., Olsson, H., & Björkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making,* 10(3), 189–209.

Ladefoged, P., & Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics, 49,* 43-51.

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication,* 30, 9–26.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17,* 249-271.

McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology, 31,* 53-65.

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2006). A Forensic Phonetic Study of ´Dynamic` Sources of Variability in Speech: The DyViS Project. *Proceedings of the 11th Australasian International Conference on Speech Science and Technology, 6-8 December 2006, Auckland: Australasian Speech Science and Technology Association, 13-18.* Collected, January 4, 2013 from www.ling.cam.ac.uk/gea/research/research.htm

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of Confidence in Earwitness Versus Eyewitness Identification. *Journal of Experimental Psychology: Applied* 1998,.4(2), 101-118

Popelka, M. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., & Klein, R. (1998). Low prevalence of hearing aid use among older adults with hearing loss: The epidemiology of hearing loss study. *Journal of American Geriatric Society*, 46, 1075–1078.

Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology, 32,* 892-905.

Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: Speaker identification.

Journal of Applied Psychology, 65, 111-116.

Solan, M., & Tiersma, P. M. (2003). Falling on deaf ears. *Legal affairs/Hastings Law Journal Jan issue.*

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology, 1,* 121-131.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part 1: Recognition of backward voices. *Journal of Phonetics, 13,* 19–38.

Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters. Part 2: Recognition of rate-altered voices. *Journal of Phonetics, 13,* 39–52.

Yarmey, A. D. (1991). Descriptions of distinctive and nondistinctive voices over time. *Journal of the Forensic Science Society, 31,* 421-428.

Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology, 6,* 367-377.

Yarmey, A. D. (1995). Earwitness speaker identification. *Public Policy and Law Psychology.* Vol.1(4), 792–816.

Yarmey, A. D. (2000). The older eyewitness. *Elders, crime, and the criminal justice system: Myth, perception, and reality in the 21st century* (pp. 127–148). New York: Springer.