Detecting Deceptive Speech: Requirements, Resources and Evaluation

Julia Hirschberg
Columbia University
New York NY 10027
USA
julia@cs.columbia.edu

Abstract

Investigations of how to detect deception in speech have lagged behind research in other types of speaker state, largely due to the lack of cleanly recorded training corpora that adequately represent the phenomenon and for which ground truth is known. We survey the current state of deceptive speech research and discuss problems and possibilities for the future.

1. Introduction

Deception is generally defined as ``a deliberate attempt to mislead others" (DePaulo 2003). Deceivers are those who attempt to convince others that something is true which the deceiver knows to be false --- thus excluding, e.g., actors or pathological liars. Distinguishing deceivers from truth-tellers is a topic of interest to scientists as well as to law enforcement personnel, who hope that scientific research will identify reliable cues which might be used by machines or humans in practical deception detection. Most deception studies today focus on visual cues to deception, such as facial expressions (e.g. Ekman 1976) or body gestures (e.g. Burgoon 1994) or on traditional biometric cues used in polygraphy (e.g. Horvath 1973). While studies associating the detection of vocal indicators of stress with deception have promised to provide simpler objective methods of detection, the earliest approaches, Voice Stress Analysis techniques, have proven disappointing (Haddad 2002, Hopkins 2005.

In recent years, there has been considerable interest in the speech community in the automatic identification of affective speech (Cowie 2003). Many current research projects are attempting to apply corpus-based machine learning approaches to the detection of emotions such as anger, frustration, confidence or uncertainty in spoken dialogue systems targeted at call centers or tutoring systems (Lee 2002, Ang 2002, Batliner 2003, Litman 2004, Liscombe 2005). Such research has motivated the application of similar techniques in attempts to identify other types of speaker state, such as deception (Hirschberg 2005, Fadden 2006, Enos 2006, Graciarena 2006), which itself has been associated in the psychological literature (Ekman 1992) with Emotions such as fear (of detection) or elation (at ***not*** being detected).

A major problem for studies of deception in any channel is the fact that many variables may influence the speaker's state during an act of deception. 'White' lies in social settings, where the consequences of detection are small, produce different psychological and physiological effects from `serious' lies, whether the stakes are high. Speakers uttering `serious' lies may also experience different emotions depending upon whether they are hiding a transgression or lying for a cause they deem worthy. Age and culture also play an important role in the feelings subjects' have about lying and thus, in the auditory and visual manifestations of their deceiving. Even within an age and cultural group, there appears to be wide variation in the cues people exhibit or recognize as indicators to deception. However, such differences, while anecdotally recognized, have as yet been little studied (but cf: Bond et al 1990 and Al-Simadi 2000).

While early studies were better able to utilize scenarios with `high stakes' deception (in which subjects could be motivated by fear or shame) in the laboratory (Mehrabian 1971), more recent studies are limited to less stressful scenarios by human subjects protocols and privacy considerations. Studies of `high stakes' lies today are limited in value since it is difficult to simulate realistic scenarios inspiring fear or true elation in the laboratory. So, most laboratory studies are conducted with subjects who are motivated to lie via financial or 'self-presentational' incentives, in which subjects are persuaded that their ability to deceive is a positive quality (DePaulo 2003). A popular scenario is the 'mock-theft' paradigm, in which subjects are given the option of taking a check or not, and then interrogated about their decision; if they can succeed in deceiving the interrogator, they are told, the check will be sent to an organization of their choice, but if they fail, they are told, it will be sent to an organization opposed to their views (Frank 1997). Other scenarios used to collect deceptive and non-deceptive data involve asking subjects to lie about the content of a movie they are watching or about the number they see on a card.

A primary difficulty of studying deception in speech is the lack of cleanly-recorded corpora of deceptive and non-deceptive speech to use for training and testing. Speech collected during earlier experiments which involved human perception studies or studies of visual cues to deception andspeech collected in the field during actual interrogations is difficult to analyze due to uniformly poor recording conditions.

Most researchers as well as practitioners would agree that there is no single cue to deception, but that multiple indicators should be sought. While few studies have focused on spoken cues, there has been considerable work on lexical and semantic indicators of deception. Generally such cues have been hand coded by trained annotators or otherwise subjectively labeled, although some simple keyword-based studies have also been conducted.

2. Perceptual and Descriptive Studies of Deception

Studies of deceptive speech and language by behavioral scientists have centered mainly on human perception of deception and descriptive analyses of deviation of syntatic or lexical usage or in pitch range or loudness when compared to some general or subject-specific 'norms' in subjects' spoken or written statements. These studies provide useful information on human perception of cues to deception and some provide correlations between human perceptions and objective measures of cues in the speech signal. However, many of the findings from previous studies have been inconclusive and even contradictory, perhaps due to variation in the motivation of the deceivers studied, to the amount of prior preparation spent devising the lie, to individual differences among speakers, or to the way in which particular features have been defined in different studies.

So, deceivers have been hypothesized to speak more than truth-tellers or to speak less (Harrison 1978, Mehrabian 1971), depending perhaps upon the care with which the lie has been prepared in advance of the telling or the desire of the deceiver to 'hold back' information. They have also been thought to exhibit more response latency or less, for similar reasons (Baskett 1974, Vrij 2000,Gozna 2004); over-rehearsed deceivers may give themselves away by answering particular questions too quickly, while under-rehearsed deceivers may need to spend more time thinking about the lie they are concocting. Deceivers have been observed to speak louder or softer when lying, to speak with higher or lower pitch (Ekman 1976b, Streeter 1977) or with a faster or slower speaking rate (Mehrabian 1971, Gozna 2004), and to exhibit more vocal tension and less vocal `pleasantness'. Studies have found that deceivers exhibit fewer disfluencies or more than truth-tellers, again perhaps depending upon the amount of rehearsal of their stories (Mehrabian 1971, Vrij 2000, Gozna 2004, Benus 2006). On similar grounds that rehearsed lies differ from normal truth-telling, deceivers are thought to make fewer admissions of forgetfulness than truth-tellers. Less well-rehearsed deceivers are said to appear less confident, to provide fewer details and scene descriptions, to be less plausible and logical in their stories, to produce more repetitions, to use more passives, negations and `indirect' speech (e.g. attributing actions and opinions to *we* or *they*, to provide fewer details, to exhibit less cognitive complexity in their speech, and to stray from the topic more frequently by mentioning peripheral events or relationships (Wiener 1968, Zuckerman 1981, Zaparniuk 1995, Vrij 2000). Many of these features are captured in various coding schemes, such as Vrij's NVB (2000) coding of non-verbal behaviors of gaze, gesture, disfluencies, response latency and speaking rate; CBCA (Criteria-Based Content Analysis),which encodes lexical content (Steller89); and RM (Reality Monitoring), which codes perceptual, cognitive, and affective information identified in subjects' statements (Vrij 2000, Masip 2005).

While some similarities have been found across studies, it is not clear how a number of these features can be objectively measured; even cues which are objectively measureable must be calibrated against a speaker-dependent baseline, which may be difficult to obtain in practice. Practitioners typically explain that they spend a good portion of an initial interview determining whether a speaker normally exhibits such behaviors as avoiding eye gaze; for these speakers, making eye contact may arouse suspicion in subsequent interrogation, while for those who do ***not*** avoid eye contact normally, gaze avoidance during interrogation might be seen as suspicious (Reid 2000). And most features involving ***what*** is said must be coded or otherwise interpreted by a human agent with some skill.

DePaulo et al.'s (2003) meta-study of cues to deception provides an excellent survey of 158 hypothesized indicators and 1338 separate estimates from previous studies. This useful study compiles results from within-subject experiments in which adult subjects were observed both lying and telling the truth, where potential cues to deception were either measured objectively in some way or were rated impressionistically by humans, in an attempt to determine which cues represent statistically significant discriminators of deceptive from non-deceptive behavior when examined across all studies which include them as factors. DePaulo examines the significance of individual cues in support of five basic hypotheses about deceivers:

1. Deceivers are less forthcoming than truth-tellers (they 'hold something back').
2. Deceivers' stories are less compelling in terms of the fluency and plausibility of their narrative; they tend to be less convincing than truth-tellers over all.
3. Deceivers appear less positive and pleasant than truth-tellers, in terms of what they say and how they say it.
4. Deceivers appear tense, due to the cognitive load of maintaining a consistent lie or to fear of discovery.
5. For similar reasons, deceivers may include more imperfections in their tales, or they may include fewer, due to prior rehearsal of what they plan to say.

While many of the cues examined in these categories are facial and body gestures, a number of possible speech and language cues are included, so it is instructive to note which of these cues are borne out across studies.

With respect to acoustic and prosodic cues to deception, DePaulo found that, across the studies examined, there was evidence of a significant difference between deceivers and truth-tellers in the proportion of overall talking time deceivers spoke vs. their conversational partner, with deceivers speaking significantly less than truth-tellers. Deception was also negatively correlated with observer impressions `verbal and vocal involvement' and with observer ratings of vocal pleasantness (cf. Burgoon 1994), while it was positively correlated with impressions of `verbal and vocal uncertainty'. Overall rater impressions of tenseness were positively correlated with deception, with both vocal tension and higher pitch being positively correlated. Note that Streeter (1977) found stronger correlations between high pitch and deception for subjects more highly motivated to deceive.

However, factors such as overall response length, length of interaction, response latency, loudness, and speaking rate, which have also been proposed as potential cues to discriminating deceptive from non-deceptive speech did not show significant differences in this meta-study. Note that (Baskett74) reports that listeners were more likely to judge speakers to be liars if they answered `too quickly' ***or*** `too slowly', which may wash out differences in this cue. Mehrabian (1971) found similar conflicting evidence for speaking rate across studies, with rate generally increasing as the speaker's comfort level increased. So these features may require more sophisticated modeling, perhaps based upon individual differences in normal production, to prove useful. Note also that Gozna (2004) found that whether subjects were seated or standing during an interview affected differences between deceptive and non-deceptive behaviors, such that speaking rate increased during deception in the standing condition but ***not*** in the seated condition, while stutters ***decreased*** only in liars who were standing; in this study response latency decreased for deceivers in both conditions. So the context of the deceptive situation appears to play an important rule in the behavioral cues deceivers exhibit.

With respect to speech disfluencies (including filled and silent pauses and hesitations), often thought to mark the speech of at least the less-rehearsed deceiver, DePaulo did not find evidence for this across studies; in fact, they found that deceivers tended to make significantly fewer 'spontaneous self corrections'. Note also more recent work on filled and silent pauses as cues to deception by Benus (2006), which shows a positive correlation between these pauses and ***truth***-telling.

Examining lexical and semantic cues to deception, as coded by human raters, DePaulo found support across studies for claims that deceivers productions are less plausible and fluent than those of truth-tellers in a number of categories hypothesized in the literature: Deceivers ***did*** provide significantly fewer details than truth-tellers and tended to make significantly more negative statements and complaints. They also tended to repeat words and phrases more often than truth-tellers did. Deceivers made fewer admissions of lack of memory and fewer expressions of self-doubt. They were significantly ***more*** likely to mention extraneous material in their speech than truth-tellers. In general, there were significant negative correlations between deception and observer ratings of the plausibility of deceivers' stories and their logical structure, and there were significantly more discrepancies and ambivalent statements in their narratives.

For other hypothesized cues to deception in this category, DePaulo's study found no significant correlations with deception. These included the proportion of unique words used by deceivers, their use of generalizing terms, self-references or mutual or group references, the use of tentative constructions (e.g. `I think'), the amount of unusual or superfluous detail they provided, their discussions of speaker or listener's mental state, the amount of sensory information they provided (coded using RM), and the cognitive complexity of their output.

However, it is important to note that even though DePaulo found no significant correlations of many hypothesized cues across the studies they included, individual studies ***have*** found these features to be useful cues to deception, either alone or in combination with other features. And more recent work has been done on some of them, which of course was not included in this meta-study. It is also difficult to combine studies of individual cues which may be subject to different definitions and interpretations, particularly when these cues are measured perceptually rather than objectively. So, while DePaulo's results are useful, they clearly do not rule out potential cues to deception.

3. Practitioners' Lore

There is also some literature and much lore among members of law enforcement agencies and the military to identify various practical auditory and lexical cues to deception for use by interviewers and interrogators. The most commonly cited oral cues for these practitioners include longer or shorter response latency, filled pauses and other disfluencies, and repetitions (Reid 2000). In most cases these cues are intended to be callibrated against a `norm' for an individual being questioned; such norms are typically established while asking interviewees questions they are likely to answer truthfully, depending upon the purpose of the interview, such as ``What is your name?" or ``What is today's date?" Considerable weight is also given to detection of deception by a close analysis of the lexical and syntactic choices of suspects' oral (transcribed) or written statements, callibrated here against a general 'normal' usage developed by practitioners over years of experience. Statement Analysis (Adams 1996) is one of the best-documented versions of this approach. Designed as a tool for interrogators, this approach looks for deviation from `normal' use of pronouns and verb tense as well as hedges (e.g. ``I think") and memory lapses in critical positions in the narratives elicited. For example, explicit use of the first person pronoun rather than a more general attribution or the absence of any subject (e.g. ``Went to the bank") is deemed normal in narrative; failure to pronominalize on subsequent mention (e.g. repeating ``My wife and I'' rather than using ``we'' is deemed abnormal). Changes in tense during a narrative, as from past to present, are also seen as suspect, indicating a place in the statement where subjects may not be telling the truth. Truthful subjects are believed to recount events chronologically and concisely, while liars will not. Such analyses must currently be performed by trained interviewers.

4. Computational Approaches to Deceptive Speech

Corpus-based, machine learning approaches to detecting deception via automatically extractable objective features have been rare, in part due to the absence of corpora recorded under suitable conditions and labeled for truth or lie. One exception is work on Voice Stress Analysis (VSA), which assumes that indicators of vocal stress also indicate deception, but this hypothesis has not been supported in experimental testing, although features examined for VSA analysis may eventually prove to be useful in combination with other features.

4.1. Lexical and Semantic Analysis

There has been some attempt to automate a simple form of lexical analysis of deceptive text in a program called Linguistic Inquiry and Word Count (LIWC), developed in the 1990s (Pennebaker 2001, Newman 2003). LIWC computes the percentage of words in a text that fall in one of 72 different categories, to capture `negative' emotion, degree of self-reference, and indicators of cognitive complexity, under the hypothesis that liars exhibit more of the first and less of the second two. Using this keyword-based analysis, (Newman03) reports classifying liars vs. truth-tellers at an overall accuracy rate of 61%.

4.2. Voice Stress Analysis

Voice Stress Analysis (VSA) approaches rely upon low level indicators of stress such as ***microtremors***, or vocal *jitter*, as indirect indicators of deception. There has been little evidence that VSA systems can effectively discriminate deception from non-deceptive speech (Haddad 2002), although (Hopkins 2005) has found that such systems might be useful tools for a skilled examiner. (Liu 2005) recently tested the utility of jitter vs. other features as discriminators for deception and found that, while jitter did ***not*** discriminate, pitch did, although only in a speaker-dependent manner. However, VSA systems continue to be marketed widely to law enforcement agencies as the answer to their deception detection problems.

4.3 Machine Learning Approaches

Recently, there has been interest in applying Machine Learning techniques to the problem of deception detection from speech, seeking to test which of the many features proposed in the behavioral literature might be a) objectively measurable and b) useful discriminators.

Qin (2004) has described preliminary studies using decisions trees trained on lexical information to predict deception. Cues included numbers of syllables, words, sentences, short sentences and 'simple' sentences; measures of word and sentence complexity; indicators of specificity and expressiveness; and an 'informality' measure based on errors that were automatically detectable. Results for the best performing decision trees examined from 20 cross-validation runs on a very small data set are reported in the mid-high 70% range.

Work has also been underway to apply speech technologies and machine learning techniques to a new, cleanly recorded corpus of deceptive speech, the Columbia-SRI-Colorado (CSC) Corpus (Hirschberg 2005, Benus 2006, Graciarena 2006). This corpus was designed to elicit within-speaker deceptive and non-deceptive speech. The corpus includes interviews with thirty-two native speakers of Standard American English. Subjects performed tasks in six areas, where the difficulty of tasks was manipulated so that interviewees scored higher than an artificial profile in two areas, lower in two, and identically in another two. Subjects received financial and self-presentational incentives to convince an interviewer that they had in fact performed the same as the target profile. Subjects were instructed to press one of two pedals hidden from the interviewer after each statement, one pedal for truth and one for lie to capture ground truth. The interviews lasted between 25 and 50 minutes, and comprised approximately 15.2 hours of dialogue; they yielded approximately 7 hours of subject speech. Data was recorded using headworn microphones in a sound booth and was subsequently orthographically transcribed. Several segmentations were created from the data: the implicit segmentation of the pedal presses, which was hand corrected to align with corresponding sets of statements; word segments, from the automatic alignment of the transcription using an SRI ASR engine; hand-labeled sentence-like units (NIST 2004); and `breath groups' which were identified from ASR word alignments plus intensity and pauses, and subsequently hand-corrected. The corpus thus consists of lexical transcription, global and local lie labels, segmentations, and the speech itself.

A series of machine learning experiments employing different learning algorithms and a variety of features sets and segmentations on this corpus has achieved classification accuracies of 66.4% (Hirschberg 2005), using a combination of acoustic-prosodic, lexical and speaker-dependent features and 64.0% using acoustic-prosodic features alone (Graciarena 2006). A human perception study performed on this data found that human judges asked to determined whether each statement was truth or lie scored on average ***worse*** than chance. Thus, the automatically produced results are quite encouraging.

5. Conclusion

The current state of deception studies from speech and language cues remains largely the domain of behavioral scientists conducting laboratory studies which peripherally include vocal cues and of practitioner proponents of various types of text-based statement analysis. Larger machine learning studies combining speech and text-based cues with potential facial, gestural, and biometric cues to deception have yet to be undertaken, largely due to the lack of corpora which include clean data from each potential cue dimension and which can be reliably labeled for truth or lie. The investigation of machine-extractable rather than hand-coded or impressionistic cues also suffers from this lack, since insufficient data for training and testing of such features is lacking. Furthermore, data used in most current research on deception are collected from subjects whose motivation for deception is probably very different from that of deceivers in the real world and scenarios closer to real life which will nonetheless be accepted by institutional review boards are hard to devise. `Real' data, collected by law enforcement agencies, is rarely recorded under conditions sufficient to do adequate acoustic-prosodic analysis, although, when transcribed, it may suffice for those focusing on lexical information -- if ground truth (was the subject ***really*** lying or not?) can be reliably established. Using such data, where it is available, also involves resolving serious ethical and legal issues. Investigation of the importance of individual and cultural differences in deception, another major area who importance is generally acknowledged, has rarely been undertaken.

In sum, the field of deception studies presents abundant open questions for research. Answering these questions, however, requires the resolution of some very difficult data collection and annotation questions, involving both technical and ethical/legal issues. It it is likely that current security concerns will provide powerful incentives for finding solutions to these issues, but it is also likely that many more `solutions' to the problem of detecting deception will be championed which have ***not*** been scientifically tested, due to the difficulty of such testing. For these reasons, it is important for behavioral scientists and speech and language technologists to work together to ensure that deception detection itself is not deceptive.

6. References

Aamodt, M. G. and Mitchell, H. (2004). Who can best detect deception: A meta-analysis. Paper presented at the annual meeting of the Society for Police and Criminal Psychology, Rome.

Adams, S. H. (1996). Statement analysis: What do suspects' words really reveal? FBI Law Enforcement Bulletin. http://purl.access.gpo.gov/GPO/LPS1534.

Al-Simadi, Fayed (2000). Social Behavior and Personality: an International Journal. 28 (5), 455-461.

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., and Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. Proceedings of the International Conference on Spoken Language Processing, pages 2037-39, Denver.

Baskett, G. D. a. R. O. F. (1974). Aspects of language pragmatics and the social perception of lying. Journal of Psycholinguistic Research, 3(2):117-30.

Batliner, A., Fischer, R., Huber, R., Spilker, J., and Noth, E. (2003). How to find trouble in communication. Speech Communication, 40:117-143.

Benus, S., Enos, F., Hirschberg, J., and Shriberg, E. (2006). Pauses in deceptive speech. Speech Prosody 2006, Dresden.

Bond, C. R, Jr., Omar, A., Mahmoud, A. & Bonser, R. N. (1990). Lie detection across cultures. Journal of Nonverbal Behavior, 14, 189- 204.

Burgoon, J. K. and Buller, D. B. (1994). Interpersonal deception: Iii. effects of deceit on perceived communication and nonverbal behavior dynamics. Journal of Nonverbal Behavior, 18(2):155-84.

Cowie, R., Douglas-Cowie, E., and Campbell, N. (2003). Speech communication: Special issue on speech and emotion.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. Psychological Bulletin, 129(1):74-118.

Ekman, P. (1992). Telling lies: clues to deceit in the marketplace, politics, and marriage. Norton, New York.

Ekman, P. and Friesen, W. V. (1976). Pictures of Facial Affect. Consulting Psychologists Press, Palo Alto, CA.

Ekman, P., Friesen, W. V., and Scherer, K. R. (1976). Body movement and voice pitch in deceptive interaction. Semiotica, 16(1):23-27.

Enos, F., Benus, S., Cautin, R. L., Graciarena, M., Hirschberg, J., and Shriberg, E. (2006). Personality factors in human deception detection: Comparing human to machine performance. Proceedings of INTERSPEECH 2006, Pittsburgh.

Fadden, L. (2006). The prosody of suspects' responses during police interviews. In Speech Prosody 2006.

Frank, M. G. and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high stake lies. Journal of Personality and Social Psychology, 72:1429-1439.

Frank, M. G. and Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. Journal of Applied Communication Research, 31(1):58-75.

Gozna, L. F. and Babooram, N. (2004). Non-traditional interviews: Deception in a simulated customs baggage Paper presented at the 14th European Conference of Psychology and Law, Krakow, Poland, July 7-10. (Also titled "The impact of a non-traditional interview setting on exhibited deceptive behaviour: A mock customs baggage search" on the first author's website.)

Graciarena, M., Shriberg, E., Stolcke, A., Enos, F., Hirschberg, J., and Kajarekar, S. (2006). Combining prosodic lexical and cepstral systems for deceptive speech detection. Proceedings of ICASSP-2006.

Haddad, D. and Ratley, R. (2002). Investigation and evaluation of voice stress analysis technology. http://www.ncjrs.org/pdffiles1/nij/193832.pdf.

Harrison, A. A., Hwalek, M., Raney, D. F., and Fritz, J. G. (1978). Cues to deception in an interview situation.

Social Psychology, 41(2):156-61.
Hirschberg, J., Benus, S., Brenier, J. M., Enos, F., Friedman, S., Gilman, S., Girand, C., Graciarena, M., Kathol, A., Michaelis, L., Pellom, B., Shriberg, E., and Stolcke, A. (2005). Distinguishing deceptive from non-deceptive speech. Proceedings of INTERSPEECH 2005, Lisbon.
Hopkins, C. S., Ratley, R. J., Benincasa, D. S., and Grieco, J. J. (2005). Evaluation of voice stress analysis technology. In Proceedings of the 38th Hawaii International Conference on System Sciences.
Horvath, F. (1973). Verbal and nonverbal clues to truth and deception during poligraph examinations. Journal of Police Science and Administration, 1:138-152.
Lee, C. M. and Narayanan, S. S. (2002). Combining acoustic and language information for emotion recognition. Proceedings of the International Conference on Spoken Language Processing, pages 873-76, Denver.
Liscombe, J., Venditti, J., and Hirschberg, J. (2005). Detecting certainness in spoken tutorial dialogues. Proceedings of INTERSPEECH 2005, Lisbon.
Litman, D. and Forbes-Riley, K. (2004). Predicting student emotions in computer-human dialogues. Proceedings of ACL-2004, Barcelona.
Liu, X. (2005).Voice stress analysis: Detection of deception.Master's Thesis at the University of Sheffield, Department of Computer Science.
Masip, J., Sporer, S. L., Garrido, E., and Herrero, C. (2005).The detection of deception with the Reality Monitoring approach: A review of the empirical evidence.Psychology, Crime, and Law, 11:99-122.
Mehrabian, A. (1971).Nonverbal betrayal of feeling. J. Experimental Research in Personality, 5:64-73.
Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. (2003). Lying words: Predicting deception from linguistic style. Personality and Social Psychology Bulletin, 29:665-675.
NIST (2004). Fall 2004 Rich Transcription (RT-04F) evaluation plan. http://www.nist.gov/speech/tests/rt/rt2004/fall/docs/rt04f-eval-plan-v14.pdf.
Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic Inquiry and Word Count. Erlbaum Publishers, Mahwah, NJ.
Qin, T., Burgoon, J. K., and Nunamaker, J. F. (2004). An exploratory study on promising cues in deception detection and application of decision tree. Proceedings of the 37th Annual Hawaii International Conference on System Sciences, pages 23-32.
Reid, J. E. and Associates (2000). The Reid Technique of Interviewing and Interrogation. Reid, John E. and Associates, Inc., Chicago.
Steller, M. and Koehnken, G. (1989). Criteria-based content analysis. In Raskin, D. C., editor, Psychological Methods in Criminal Investigation and Evidence, pages 217-245. Springer-Verlag, New York.
Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., and Apple, W. (1977). Pitch changes during attempted deception. Journal of Personality and Social Psychology, 35(5):345-350.
Vrij, A. (1994). The impact of information and setting on detection of deception by police detectives. Journal of Nonverbal Behavior, 18(2):117-136.
Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. Journal of Nonverbal Behavior, 24(4):239-63.
Wiener, M. and Mehrabian, A. (1968). Language within Language: Immediacy, a Channel in Verbal Communication. Appleton-Century-Crofts, New York.
Zaparniuk, J., Yuille, J. C., and Taylor, S. (1995). Assessing the credibility of true and false statements. International Journal of Law ans Psychiatry, 18:343-352.
Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). Verbal and nonverbal communication of deception.In Berkowitz, L., editor, Advances in Experimental Social Psychology, volume 14, pages 1-59. Academic Press, New York.