Automatic Detection of Verbal Deception

Synthesis Lectures on Human Language Technologies

Editor
Graeme Hirst, ***University of Toronto***

Synthesis Lectures on Human Language Technologies is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

Automatic Detection of Verbal Deception
Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
2015

Semantic Similarity from Natural Language and Ontology Analysis
Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain
2015

Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition
Hang Li
2014

Ontology-Based Interpretation of Natural Language
Philipp Cimiano, Christina Unger, and John McCrae
2014

Automated Grammatical Error Detection for Language Learners, Second Edition
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2014

Web Corpus Construction
Roland Schäfer and Felix Bildhauer
2013

Recognizing Textual Entailment: Models and Applications
Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto
2013

Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax
Emily M. Bender
2013

Semi-Supervised Learning and Domain Adaptation in Natural Language Processing
Anders Søgaard
2013

Semantic Relations Between Nominals
Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz
2013

Computational Modeling of Narrative
Inderjeet Mani
2012

Natural Language Processing for Historical Texts
Michael Piotrowski
2012

Sentiment Analysis and Opinion Mining
Bing Liu
2012

Discourse Processing
Manfred Stede
2011

Bitext Alignment
Jörg Tiedemann
2011

Linguistic Structure Prediction
Noah A. Smith
2011

Learning to Rank for Information Retrieval and Natural Language Processing
Hang Li
2011

Computational Modeling of Human Language Acquisition
Afra Alishahi
2010

Introduction to Arabic Natural Language Processing
Nizar Y. Habash
2010

Cross-Language Information Retrieval
Jian-Yun Nie
2010

Automated Grammatical Error Detection for Language Learners
Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault
2010

Data-Intensive Text Processing with MapReduce
Jimmy Lin and Chris Dyer
2010

Semantic Role Labeling
Martha Palmer, Daniel Gildea, and Nianwen Xue
2010

Spoken Dialogue Systems
Kristiina Jokinen and Michael McTear
2009

Introduction to Chinese Natural Language Processing
Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang
2009

Introduction to Linguistic Annotation and Text Analytics
Graham Wilcock
2009

Dependency Parsing
Sandra Kübler, Ryan McDonald, and Joakim Nivre
2009

Statistical Language Models for Information Retrieval
ChengXiang Zhai
2008

Copyright © 2015 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Automatic Detection of Verbal Deception

Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari

www.morganclaypool.com

ISBN: 9781627053372 paperback
ISBN: 9781627053389 ebook

DOI 10.2200/S00656ED1V01Y201507HLT029

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #29
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Print 1947-4040 Electronic 1947-4059

Automatic Detection of Verbal Deception

Eileen Fitzpatrick
Montclair State University

Joan Bachenko
Linguistech LLC

Tommaso Fornaciari
Italian National Police

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #29

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

The attempt to spot deception through its correlates in human behavior has a long history. Until recently, these efforts have concentrated on identifying individual "cues" that might occur with deception. However, with the advent of computational means to analyze language and other human behavior, we now have the ability to determine whether there are consistent clusters of differences in behavior that might be associated with a false statement as opposed to a true one. While its focus is on verbal behavior, this book describes a range of behaviors—physiological, gestural as well as verbal—that have been proposed as indicators of deception. An overview of the primary psychological and cognitive theories that have been offered as explanations of deceptive behaviors gives context for the description of specific behaviors. The book also addresses the differences between data collected in a laboratory and "real-world" data with respect to the emotional and cognitive state of the liar. It discusses sources of real-world data and problematic issues in its collection and identifies the primary areas in which applied studies based on real-world data are critical, including police, security, border crossing, customs, and asylum interviews; congressional hearings; financial reporting; legal depositions; human resource evaluation; predatory communications that include Internet scams, identity theft, and fraud; and false product reviews. Having established the background, this book concentrates on computational analyses of deceptive verbal behavior that have enabled the field of deception studies to move from individual cues to overall differences in behavior. The computational work is organized around the features used for classification from *n*-gram through syntax to predicate-argument and rhetorical structure. The book concludes with a set of open questions that the computational work has generated.

KEYWORDS

credibility assessment, deception detection, factual language, forensic linguistics, gold-standard data, ground truth, high-stakes scenarios, imaginative language, real-world data, stylometry, text classification

Eileen Fitzpatrick: To my husband, Ralph Grishman

Joan Bachenko: In memory of my mother, Claire Joan Baumgartner

Tommaso Fornaciari: In memory of my father, Alberto Fornaciari

Contents

Preface

There are many venues where the ability to spot the lie is an important, often critical, skill. It is necessary in police, security, border crossing, customs, and asylum interviews; in congressional hearings; in financial reporting; in legal depositions; in human resource evaluation; in predatory communications, including Internet scams, identity theft, and fraud; and in false advertising. Discovering lies can thwart serious immediate threats, provide productive directions in the investigation of past events, and assist in the accurate prediction of likely future events.

For much of the 20th century, the fields of psychology and criminal justice have studied the behaviors that might be associated, directly or indirectly, with deception. Three types of behavior have been examined: physiological behavior; vocal behavior, including prosodic features; and verbal behavior, including the words and structures that might correlate with deception.

The study of verbal behavior in deception is relatively new and the attention that natural language processing has paid to discriminating true from false claims is even newer, with most of the work done in the last 10 years as classification techniques have improved. Now is a good time to review the prior literature on deception and consider the NLP approaches that have been tried. Knowing the foundations and trends in work on deception, both theoretical and applied, will enable us to move forward productively.

Several areas of NLP are ripe to address the vocal and verbal features that might be associated with deception and new approaches may well combine information from these with the facial and body movements associated with deception. A spate of recent NLP papers on the classification of narratives as truthful or deceptive suggests that the field is ready to open up to this promising area.

The genesis of this text was a workshop on deception detection that took place as part of the European Meeting of the Association for Computational Linguistics in Avignon in the spring of 2012. The workshop brought together 35 colleagues and offered 14 presentations on work that ranged from annotation tools and corpus building for deception to cross-linguistic classification of deceptive narrative. It also gave the authors of this text an opportunity to work together on our common interest in the use of "real-world" data, primarily legal data, in NLP deception studies.

Following the mission of the Synthesis series, the present text is designed to give the student or researcher in natural language processing a background in the history of deception studies concentrating on the behavioral cues to deception that have been supported in the psychology, applied psychology, and criminal justice literature, a consideration of the real-world data sources for NLP work in deception, a review of NLP work in deception organized around the features

used for classification from ngram to rhetorical structure, and a look at exciting questions and areas that need to be addressed in order for the field to progress.

Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
July 2015

Acknowledgments

We wish to thank the students who took part in a graduate seminar on current issues in NLP at Montclair State University in which a previous draft of this text was used as the primary text for the course, including Richard Barrett, Janette Martinez, Matthew Mulholland, and Emily Olshefski. Their ability to take the material covered in the text to the next step in their own research gives us confidence that we are respecting the mission of this series.

We would also like to thank Ralph Grishman and Massimo Poesio for input, checking, and fruitful discussions related to the issues in this book.

We are indebted to Myle Ott and another, anonymous, reviewer who provided detailed, thoughtful feedback in their reviews of a draft of the book, and to the series editor Graeme Hirst who first suggested a text on deception in Avignon and has guided it through, with great patience and care, to reality. We are also grateful to the many authors whose work we have cited here.

Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari
July 2015

CHAPTER 1

Introduction

1.1 INTRODUCTION

Deception occurs frequently in everyday situations, from insincere compliments — "You look great!" — to face saving lies "Can you lend me $5? I lost my ATM card." Most of these lies are inconsequential; some even have positive effects. This book concentrates on the more consequential lies and on the behaviors that are thought to be associated with lying, particularly lying involving natural language. This book discusses how these behaviors are being captured by applications in natural language processing.

The study of deception has deep implications for human, and some animal, behavior since it demonstrates an awareness of self, in particular an awareness that one's own thoughts can differ from those of others [Keenan et al., 2005].

In the broadest sense, deception includes self-deception, acting, and conjuring. It also sometimes covers false statements believed by the teller to be true. This book, however, is devoted to applications where something important is at stake in communication: in police, security, border crossing, customs, and asylum interviews; in congressional hearings; in financial reporting; in legal depositions; in human resource evaluation; in predatory communications, including Internet scams, identity theft, and fraud; and in false advertising. For the purposes of this book, then, we will follow the definition of deception given by Vrij (2008), which excludes self-deception, acting, and falsely held beliefs. Vrij defines deception as a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue.

The liar can carry out the deception in different ways. The way that immediately comes to mind is the outright lie ("I was not required to approve those transactions"),[1] but liars can be evasive ("Well, there's an issue as to whether I was actually at a—the particular meeting that you're talking about"), exaggerate or minimize an issue ("In that meeting, the power had gone out, and as everybody remembers,...the room was dark, quite frankly, and people were walking in and out of the meeting") or omit significant facts from a story, as did Scott Peterson, convicted for the murder of his pregnant wife, in his detailed account of his actions on the day she was reported missing.

The attempt to spot deception has a long history, dating at least from the Greek physician Erasistratus (300–250 B.C.), who felt the pulse of a suspect to distinguish the lie from the truth

[1]The deceptive quotes in this paragraph and the next are from the testimony of Jeffrey Skilling, CEO of Enron, the American energy company, to Congress on February 7, 2002 concerning the accounting fraud that his company had perpetrated.

[Trovillo, 1939], still a measure used in modern polygraphy. While a good deal of research has been devoted to potential physiological cues to deception since the polygraph was invented in 1921, including imaging technologies such as functional magnetic resonance imaging (fMRI) and thermal imaging, visual measures that include body movements, and vocal measures such as pitch and speech rate changes, there is an equally robust literature on the characteristics of language that are thought to be associated with deception. In the psychology and criminal justice literature on deception, these range from discrete cues such as higher rates of negative statements and extreme descriptions ("absolutely, positively no connection") to more global features like lexical diversity and story consistency. To give the scope of work in deception detection, we provide in Chapter 2 an overview of both the non-verbal and verbal studies on deceptive behavior.

The computational literature, which has, to a great extent, replicated the findings of the psychological experimentation, runs the gamut of comparisons of true versus false statements from differences in n-gram occurrences to differences in local features of the narrative to distinctions in rhetorical structure. In its establishment of a baseline against which to measure success, its use of classification algorithms to separate true from false narratives, its training and testing procedures, and its reliance on standard evaluation measures to estimate success, the computational work has the characteristics of much work in applied natural language processing. Unlike many NLP ventures, though, it is hampered by data limitations: it is hard to come by narratives the proposition(s) of which are known to be true or false. We devote Chapter 3 to attempted solutions to this problem.

In addition, computational studies of verbal deception differ from work in other areas of natural language processing in two important, related respects: most NLP studies regard human performance as the gold standard, whereas typical human performance in detecting deception runs at chance levels. This first difference leads to a second: while most evaluations in NLP test against human performance, work in deception detection must test against ground truth, which is external to the verbal data—Was, for example, Enron's Jeffrey Skilling present at that critical meeting or was he not?

1.2 VERBAL CUES TO DECEPTION

Propositional communication takes place through words, and so does the opportunity to misrepresent reality. The most obvious way to detect deception in communication, then, would be to compare the propositions with the reality: if some dissimilarity is found, and the narrator's awareness of that dissimilarity is (reasonably) demonstrated, the communication can be regarded as deceptive.

When doubts about the truthfulness of a communication arise, typically a great deal of effort is directed at establishing the ground truth; getting at this truth is one of the main goals not only in court trials and in police investigations, but also in private disputes. Unfortunately, determining the ground truth in most cases is difficult if not impossible. In such cases, having

robust cues to deception would at least lead the inquirer toward communications that are more likely to be lies.

A notable example of the use of linguistic analysis to uncover ground truth dates back to the 15th century, when the Italian rhetorician Lorenzo Valla proved the forgery of the Donation of Constantine, the fake imperial decree which supposedly conferred on the Pope authority over the western Roman Empire. In that case, Valla demonstrated the falsity of the document not only by discussing the historical implausibility of the Donation and addressing a clear mistake with respect to the Donation's date, which referred to a time not compatible with the content of the document, but also by making use of linguistic arguments. In particular, he emphasized that the Latin of the Donation could not belong to the imperial period, but was typical of the following centuries [Valla, 2008].

The study of verbal cues to deception assumes that the style of the communication is affected not only by the demographic, social, and cultural characteristics of the narrator, but also by his state of mind at the moment of the production of the communication. In particular, the basic assumptions of this approach are as follows:

- the psychological condition of the subject affects communication style;
- the elaboration of a lie and the recall of a memory are different cognitive processes; and
- at least in high stakes scenarios, the emotional charge of a lie differs from that of a truthful statement.

It assumes, then, that the communication style of a liar may be different from that of a truth teller. Basically, this is the same idea that characterizes the studies of non-verbal behavior and physiological variables with respect to deception. However, while these studies belong to the research fields of psychology and physiology, the theoretical paradigm here comes from linguistics, and in particular from the structuralist school of thought. The structuralist approach studies texts through the relations existing among their single elements: this mode of reasoning gave rise to the possibility of quantitative and computational analyses of the texts.

The scientific study of deception in language dates at least from Undeutsch, who hypothesized that it is "not the veracity of the reporting person but the truthfulness of the statement that matters and there are certain relatively exact, definable, descriptive criteria that form a key tool for the determination of the truthfulness of statements" [Undeutsch, 1954].

In the last ten years, modern natural language processing techniques have been applied by many researchers to the detection of deception with promising results. This research is covered in Chapter 4. The achievement of these studies is not insignificant, since identifying deception by any means has proven to be a very difficult task, regardless of the kind of cues employed to elicit the deceit. As Aldert Vrij, who has carried out extensive studies of deception within social psychology, notes in referring to the lying protagonist of Italian fable: "a verbal cue uniquely related to deception, akin to Pinocchio's growing nose, does not exist" [Vrij, 2008, p. 103].

Given this situation, the fair success of NLP analyses carried out through machine learning techniques is probably due, at least in part, to the fact that this line of research relies on clusters of cues, which provide an overall picture of the deceptive language. Nevertheless, the effectiveness of machine learning techniques depends equally on the effectiveness of the individual cues, or features, of deception that distinguish the narratives under analysis. Therefore it is worth looking closely at the linguistic features employed in detecting deception in communication.

1.2.1 LINGUISTIC FEATURES USED IN IDENTIFYING DECEPTION

The goal here is to distinguish truthful (T) from false (F) narratives with a high degree of success. To that end, NLP research has borrowed cues to false statements from the psychology and criminal justice literature as well as from standard techniques in natural language processing.

The standard NLP approach uses surface level elements of the narrative—characters, *n*-grams, part-of-speech tags, narrative length, lexical diversity—in a pure classification task. The paucity of data in this field constrains the use of other elements commonly used in NLP, for example, collocational properties.

The data constraints have driven many investigators to generalize over semantically related linguistic elements, for example, self-referencing items (*I, me, my*) as opposed to items that reference others (*you, they, them*). This approach also has the advantage of tapping into psychological motivation for these features; self-references have been found to appear much less frequently in deceptive narratives than in truthful ones, which makes sense if one is trying to distance oneself from the narrative. The most commonly used features here are those of the Linguistic Inquiry and Word Count software [Pennebaker et al., 2007].

In a similar vein, studies have devised objective ways of characterizing certain of the findings from the psychology literature. Verbal and vocal immediacy, for example, are identified by many studies as highly discriminating between T and F narratives as indicated by DePaulo et al. [2003] and have been measured in the NLP literature by presence of active voice, present tense, and self-referencing.

More recently, deeper and/or more global characteristics of the narrative are being investigated to enhance the classification performance. Among these are parse tree differences between T and F narratives. The collection of syntax-related features is more complex than that of surface features, since it requires parsing the narrative in order to identify its syntactic structures, encoded as trees—or parts of trees—of syntactic elements which are used as features of the texts. In the field of deception detection, this approach was followed by Feng et al. [2012a], who found that performance in the classification of narratives as truthful or deceptive was notably better when deep syntactic features were employed instead of shallow syntactic features such as part of speech.

Narrative coherence and discourse relations within the narrative are also under investigation by Rubin and Vashchilko [2012]. This is consistent with work by Susan Adams within the criminal justice field on person-of-interest narratives written prior to police interviews, which notes an imbalance in deceptive narratives among the introduction, body, and conclusion [Adams, 1996,

2002]. Another promising approach creates a profile representative of T narratives against which a test narrative can be measured for compatibility [Feng and Hirst, 2013]. Chapter 4 is devoted to greater coverage of the NLP work on deception.

1.2.2 EFFECTIVENESS OF LINGUISTIC CUES TO DECEPTION

We will see that the variation in the success of T/F classification within NLP depends largely on the contextual factors involved in each study, including the topics, genres, registers, and modes of delivery (face-to-face, written, electronic, etc.). Depending on these factors, classification accuracy rates vary from the low 90% range to the low 60% range, which is still better than the random accuracy of most human judgments.

If we want to achieve classification systems that can generalize across these factors, the field needs to aim for a common data set and shared task against which to test the systems while still identifying the tasks that are more amenable to classification. Several groups are concerned with these issues, including Myle Ott and his colleagues, as well as Fornaciari and Poesio, Rubin and Conroy, and Fitzpatrick and Bachenko. We will consider the issues they raise in Chapters 3 and 5.

As discussed in the previous subsection, a quite wide variety of linguistic variables can be used as deception indicators. Even though none of them can be considered a high probability indicator of deception like Pinocchio's nose, the results of the studies mentioned above suggest that their combination can be useful in identifying deceptive communications.

However, while there have been a large number of studies concerning nonverbal cues to deception, which were even the main focus of a well-known American television series *Lie to Me*, there has been only a relatively small scientific community of linguists, psychologists, and computer scientists dealing with verbal cues to deception. Given that many aspects of speech elude the conscious control of the narrator, such as the aforementioned vocabulary richness, the study of verbal cues to the lie promises to provide valuable support in the difficult task of identifying deception in communication. The state of the art in this field, particularly in its automation, is the object of this text.

1.2.3 VERBAL CUES TO GROUND TRUTH

For studies that use real-world data, the establishment of what is referred to as 'ground truth' usually involves the comparison of a proposition with external data. Jeffrey Skilling, for example, indicated that he may not have even attended a critical meeting until the minutes of the meeting showing Skilling as a participant were produced. However, ground truth can also be verified by attributes of the verbal narrative itself, including the following.

Consistency involves the repetition of the same content in different statements (issued by the same or different subjects).

Contradiction involves two claims the facts of which are at odds with each other.

To make a decision about the truthfulness of a proposition involves the use of the rules of logic, pragmatics, and probability calculus. Even though the modern formalization of these concepts is quite recent, historically the application of these tools to the detection of deception is ancient, and testimonies can be found even in the Bible, for example, in the Book of Daniel (2nd century BCE), where the episode of Susana is described. Here the prophet Daniel unmasks the deceptive accusations of two old judges against the woman (Daniel 13:1-59 Nova Vulgata) by identifying an inconsistency in the different statements that he asked the two judges to issue separately, regarding the same details.

1.3 WHAT'S AHEAD

This book is designed to give someone with an introductory background in natural language processing and/or machine learning an understanding of the current approaches to the automatic detection of verbal deception. This subfield of NLP is in its infancy and so presents an exciting area in which to do groundbreaking work. The purpose of this book is to equip the reader with the information to do that.

Much of the current research in the broader field of deception detection is based on experimental work that attempts to connect specific behaviors with lying. Some work tests connections between physiological measures and lying, while the rest examines behavioral cues tested within the fields of psychology and criminal justice. We begin in Chapter 2 with a brief review of the literature on physiological cues to deception followed by a longer review of the psychology literature based on two overarching works by Bella DePaulo and her colleagues, DePaulo et al. [2003] and Aldert Vrij [2008], as well as work in applied psychology and criminal justice. Chapter 3 deals with sources of deceptive verbal behavior, primarily in the "real world." The heart of the book, Chapter 4, considers issues involved in designing an NLP experiment to test a deception system and examines the current systems that have been built to detect deception, comparing the methods and the results of these systems. Chapter 5 considers open research questions and future directions.

CHAPTER 2

The Background Literature on Behavioral Cues to Deception

2.1 INTRODUCTION

As mentioned in Chapter 1, there is a long history of attempts to link lying to measurable effects on the liar. The effects considered have been physiological in nature—for example, changes in blood pressure or vocal pitch. They have also been cognitive, as are speech disfluencies and repetitions. Emotional effects such as negative affect and verbal uncertainty have also been examined. Both the cognitive and emotional effects can be connected directly to language, as are the examples given here.

The physiological effects are at some remove from language, though vocal changes are more closely connected to it. We include them here to provide a background to the psychology and criminal justice literature, which examines all three types of effects. There are also current attempts to link the verbal effects with the physiological, which we discuss briefly in Section 5.8 of Chapter 5. Finally, the difficulties connecting physiological behaviors to deception demonstrate that the problem of identifying the lie, by any means, is far from solved.

As for the cognitive and emotional effects, there is a rich tradition, dating from Undeutsch, of the study of deceptive behaviors in experimental psychology, where data is obtained by experimentation with subjects in laboratory settings. Another, more recent, thread of studies of behaviors linked to deception comes from the applied psychology and criminal justice literature, where data is collected post hoc from police interviews, court testimony, interviews, and the like. The ramifications of each type of data collection are discussed in Chapter 3; here we examine the literature in the three traditions, reviewing the types of cues, with an emphasis on verbal cues, that have been studied.

2.2 NONVERBAL CUES TO DECEPTION

Nonverbal cues occur independent of language. They include physiological activity, vocalizations, and movements of the face and body. The list in Table 2.1, while not exhaustive, identifies the primary nonverbal cues cited in the lie detection literature and the primary manner of detection for each.[1]

[1]See text for references on results greater than chance.

Table 2.1: List of primary nonverbal cues

Cue Type		Cue Detection	Physical Contact	Results >chance	Commercialized
physiological	respiration, electro-dermal activity, blood pressure	polygraph	yes	yes	yes
	laryngeal frequencies	voice stress analyzer	no	no	yes
	unknown (proprietary)	layered voice analysis	no	no	yes
	facial blood flow	thermography	no	yes	no
	electrical brain waves	EEG	yes	yes	yes
	brain blood flow	fMRI	yes	yes	yes
vocal	voice f0, filled pause, silent pause, disfluencies, etc.	pitch analyzer, speech editor, manual analysis	no	yes	no
face/body movements	micro-expressions, pupil dilation, finger tapping, etc.	video recorder, manual analysis	no	yes	no

Of the cues in this list, physiological indicators are perhaps the most familiar because of their reliance on specialized technologies—polygraph, fMRI, electroencephalography (EEG), etc.—that have allowed the development of systems currently used in threat assessment, criminal investigation, and federal employee screening. Our review begins with this class of nonverbal cues.

Physiological cues and technologies fall into four main categories: polygraphy, voice analysis, facial thermography and brain scans. All assume that lying is a stressful activity that triggers measurable changes in the activity of some physiological system. With the exception of voice stress and layered voice analysis, the proposed physiological cues and related technologies provide at least weak support for this idea.

2.2.1 POLYGRAPHY

Polygraphy is the oldest and best established technology for associating physiological activity with deception-induced stress. Examinees are attached to at least three kinds of physiological data sensors: blood pressure cuff, electrodermal sensor, and respiration sensors positioned on the chest and abdomen [American Polygraph Association, 2010]. The test itself is usually embedded in a longer interview that may last for as long as four hours. At certain points in the interview the polygrapher will ask a series of yes/no questions. Some of these are intended to elicit physiological states that provide baseline measurements, others are intended to elicit departures from the baseline that indicate an emotionally aroused state. Aroused states presumably encode a flight instinct indicative of deception.

Evaluating the polygraph's performance depends on issues that have little to do with polygraph technology. A commonly noted complication is that the physiological states that may indicate deception often arise when deception is absent [National Research Council, 2003, Saxe et al., 1985]. In addition, test outcomes, measured by success in identifying deceptive and truthful subjects, depend largely on the skill of the interviewer, who uses the polygraph as an interrogation tool, and on characteristics of the interviewee, who may be suggestive, anxious, and inexperienced [Saxe et al., 1985]. Finally, polygraphs are well known to be vulnerable to countermeasures, techniques the interviewee can use to deliberately alter physiological states, making it possible for a deceptive interviewee to appear truthful.

Vrij [2008] and the National Research Council Report [2003] raise another concern: standardized methods for representing and scoring polygraph data are difficult to formulate and have yet to be developed. Hence there is no consistent way to tell if failure or success of a polygraph test is due to physiological measurement or to the impressions and experience of the interviewer. Not surprisingly, reports of polygraph accuracy vary widely. The review of polygraph studies by Saxe et al. [1985] cites results of field studies in which correct guilty decisions ranged from 70.6–98.6% and correct innocent decisions ranged from 12.5–94%. The NRC report concludes: "in populations of examinees such as those represented in the polygraph research literature, untrained in countermeasures, specific-incident polygraph tests for event-specific investigations can discrimi-

nate lying from truth telling at rates well above chance, though well below perfection" [National Research Council, 2003, p. 214].

Despite these criticisms, the absence of other viable alternatives makes the polygraph a widely used technique for detecting deception. Recent changes in standards of evidence have led several states to admit polygraph results into evidence [American Polygraph Association, 2010].

2.2.2 VOICE ANALYSIS: VSA AND LVA

Voice analysis cues depend on detectable frequencies produced by the body during speech. Two competing methods have been implemented for lie detection: Voice Stress Analysis (VSA) and Layered Voice Analysis (LVA). Both are available as commercial products that are popular with law enforcement professionals.

VSA technology is based on the theory that all muscles in the body, including those of the larynx, vibrate at a rate of 8–12 Hz [Lippold, 1971]. These inaudible microtremors are suppressed when a speaker experiences stress. VSA specialists and vendors claim that their technology is capable of detecting and measuring variations in laryngeal microtremor frequencies. They further claim that these variations are associated with aroused states that indicate deception.

A VSA machine is essentially a computer with VSA software that ostensibly records laryngeal microtremor patterns. The VSA machine may be used in a real-time interview or it may process pre-recorded speech. Several researchers have evaluated VSA devices in laboratory experiments [Haddad et al., 2001, Horvath, 1982] and field tests [Damphousse, 2008]. These studies have failed to confirm that microtremors exist or that VSA technologies can detect them, although there is some agreement with Haddad's [2001, p. 11] conclusion that VSAs are measuring something, but not microtremors. Moreover, despite VSA's popularity in law enforcement organizations, the tests of VSA systems have failed to show that VSA devices perform at a level above chance.

Layered Voice Analysis (LVA) is developed and marketed by Nemesysco. LVA does not use laryngeal microtremors but relies instead on an undocumented signal processing algorithm that employs a "proprietary set of vocal parameters ... new to the world of phonetics" (`www.nemesysco.com`). The description of LVA technology is too inadequate to support evaluation of its theoretical basis. This leaves performance evaluations, which have failed to provide evidence that LVA performs better than chance in laboratory tests [Harnsberger et al., 2009] and field trials [Horvath et al., 2013]. Despite the poor performance results, law enforcement professionals have reported great success in using LVA and VSA machines to solve crimes [Haddad et al., 2001]. [Horvath et al., 2013, p. 390] speculate that the reported success by field practitioners comes not from the value of the LVA, but rather from operators' ability to "read" the cues inherent in an interviewee's behavior: their tone of voice, assertiveness, directness, naturalness, and so forth. In other words, as with the polygraph, these devices succeed not on their own but only when used as supporting tools in the hands of a skillful and experienced interviewer.

2.2.3 THERMOGRAPHY

Thermal imaging works by using heat detecting cameras to identify warming patterns around a subject's eyes. Warming patterns are formed as a physiological response to stress: blood flow to the area around the eyes is increased, creating increased warmth. Pavlidis et al. [2002] claimed that it is possible to identify a "thermal signature" consisting of blood flow patterns indicative of deception and that these patterns could be used to identify deceptive subjects with "an accuracy comparable to that of polygraph examination." The appeal of this approach is that it offers the possibility of identifying deception without the need for interviews or physical contact. Hence, it would seem to hold great promise for airport and border crossing applications as suggested by Warmelink et al. [2011] and Vrij et al. [2010].

Laboratory studies of thermal imaging show some support for facial thermography patterns as an indicator of deception. In the 2011 airport study by Warmelink, thermal imaging managed to identify liars 69% of the time and truthtellers 64% of the time, a rate that they claim is too low for airport screening, especially given that interviewers working without thermography performed significantly better on the discrimination task. Results of tests by Pollina et al. [2006] suggest a stronger link between facial heat displays and deception. They conclude, however, that the status of thermography remains unclear: "The extent to which thermography will increase accuracy beyond that which is possible using traditional polygraph measures is not yet known" (p. 1189).

2.2.4 BRAIN SCAN: EEG AND MRI

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are well-established technologies for the measurement of brain activity. Their relevance to deception analysis lies in the demonstrated ability of EEG and fMRI measurements to distinguish the brain's responses to known information from responses to novel information. Hence, these technologies could be used to determine, for example, whether a suspect has special knowledge of a crime that only a guilty person would have. There exists an extensive research literature on memory and EEG/fMRI as well as significant commercial development in the United States (`www.brainwavescience.com`) and India (`www.axxonet.com`).

EEGs record brain waves called event related potentials (ERPs). The ERP memory recognition response is the P300, so named because the response usually occurs 300–900 ms after information relating to the memory is presented. The P300 response fails to occur if the information is unfamiliar and hence may be viewed as a diagnostic to determine if a subject has experiential knowledge of some event.

The measurement of P300 responses forms the heart of commercialized EEG/P300 systems that claim to detect memories indicating a suspect's guilt. Two of the strongest challenges to EEG/P300 systems come from studies of false memories and countermeasures. Allen and Mertens [2009] found that subjects' ERP responses failed to show a distinction between true recollections and false recollections that were implanted by association with true memories, opening up the possibility of deeming an innocent person guilty. In a study of countermeasures, Bergström