Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 1–21

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features ☆

Matthew P. Black [a,*], Athanasios Katsamanis [a], Brian R. Baucom [b], Chi-Chun Lee [a], Adam C. Lammert [a], Andrew Christensen [c], Panayiotis G. Georgiou [a], Shrikanth S. Narayanan [a,b]

[a] *Signal Analysis & Interpretation Laboratory (SAIL), University of Southern California, 3710 McClintock Ave., Los Angeles, CA 90089, USA*[1]

[b] *Department of Psychology, University of Southern California (USC), 3620 McClintock Ave., Los Angeles, CA 90089, USA*

[c] *Department of Psychology, University of California, Los Angeles (UCLA), 1285 Franz Hall, Los Angeles, CA 90095, USA*

Available online 16 December 2011

Abstract

Observational methods are fundamental to the study of human behavior in the behavioral sciences. For example, in the context of research on intimate relationships, psychologists' hypotheses are often empirically tested by video recording interactions of couples and manually coding relevant behaviors using standardized coding systems. This coding process can be time-consuming, and the resulting coded data may have a high degree of variability because of a number of factors (e.g., inter-evaluator differences). These challenges provide an opportunity to employ engineering methods to aid in automatically coding human behavioral data. In this work, we analyzed a large corpus of married couples' problem-solving interactions. Each spouse was manually coded with multiple session-level behavioral observations (e.g., level of blame toward other spouse), and we used acoustic speech features to automatically classify extreme instances for six selected codes (e.g., "low" vs. "high" blame). Specifically, we extracted prosodic, spectral, and voice quality features to capture global acoustic properties for each spouse and trained gender-specific and gender-independent classifiers. The best overall automatic system correctly classified 74.1% of the instances, an improvement of 3.95% absolute (5.63% relative) over our previously reported best results. We compare performance for the various factors: across codes, gender, classifier type, and feature type.

Keywords: Behavioral signal processing (BSP); Couple therapy; Dyadic interaction; Human behavior analysis; Prosody; Emotion recognition

1. Introduction

In psychology and psychiatry, behavioral observation is essential for diagnosis for children and adults, and it is also a means for monitoring change during psychotherapy, where both therapist and client engage in, and respond to, continuous, albeit usually unsystematic, behavioral observation. The importance of observable behavior for researchers and therapists is borne of the fact that behavior is typically the best objective measure of psychologically relevant phenomena available. Self-reports of even obvious behaviors can be notoriously unreliable (O'Brien et al., 1994).

Although most observation in psychological and psychiatric practice has been unsystematic, systematic observational research has been central to numerous intra- and interpersonal psychological problem domains including depression (Baucom et al., 2007), bi-polar disorder (Fredman et al., 2008), anxiety (Beck et al., 2006), schizophrenia (Brüne et al., 2008), autism (Keen, 2005), alcoholism (Shoham et al., 1998), domestic aggression (Margolin et al., 2004), and marital distress (Heyman, 2001). In each of these areas, observational research has identified behaviors

☆ "The original version of this paper was selected as one of the best papers from Interspeech 2010. It is presented here in revised form following additional peer review".

* Corresponding author. Tel.: +1 213 740 6432; fax: +1 213 740 4651. *E-mail address:* matthepb@usc.edu (M.P. Black).

[1] The SAIL homepage is http://sail.usc.edu.

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.12.003

exhibited by individuals who suffer from such problems (and behaviors exhibited by family members and loved ones of afflicted individuals) that are associated with increased symptomatology and reoccurrence of disorders.

Behavioral observation has been used with considerable success in the study and treatment of intimate relationships. Current theory suggests, and recent empirical findings validate (Karney and Bradbury, 1995; Gonzaga et al., 2007), that spouses' behavior is a central and defining aspect of intimate relationships that links broad cultural factors, longstanding life experiences, and current stressors to the stability and quality of marital relationships.

However, the methods used in behavioral observation do present some challenges. To test research hypotheses, psychology and other fields in the behavioral sciences oftentimes rely heavily upon observational coding of audio-video data; for example, in family studies research, psychologists use a variety of established coding standards describing characterizations of specific behavior patterns of interest that guide human annotation of data (Margolin et al., 1998). This manual coding is a costly and challenging process. First, a detailed coding manual must be designed, which can be a complex iterative task (Kerig and Baucom, 2004).

After the creation of an appropriate coding manual, multiple coders, each with his/her own biases and limitations, must be trained in a consistent manner on held-out but representative data. In some cases, coders must meet a predetermined minimum level of agreement with a "gold-standard" coder on training data before they can code real data. To avoid coder drift, some coding protocols require coders to be evaluated periodically and retrained if necessary (Kerig and Baucom, 2004). In addition, for longitudinal studies lasting several years, it is usually only feasible to have disjoint sets of coders, which adds another source of variability to the resulting coded data.

The actual coding process can be mentally straining and inefficient. Multiple coders oftentimes code the same data to allow for the computation of both code reliability and inter-rater reliability. Each coder observes the audio-video data and marks relevant behavioral phenomena according to the coding manual (e.g., in continuous time, in quantized time intervals, at the session-level). The complexity of the coding process determines the speed at which data can be coded, with more complex protocols taking orders of magnitude longer than real-time (e.g., Hops et al., 1971). To prevent evaluator fatigue, coders are often limited to coding for short periods of time in one sitting. Overall, the coding process is limited by the inherent subjective and qualitative nature of human descriptions of human behavior.

Technology has the potential to aid in coding human behavioral data. Computers are better suited to track and quantify certain behavioral phenomena that may be challenging, or even impossible, for humans to do. For example, whereas a human observer might have a qualitative idea of how a speaker's pitch may be changing, engineering algorithms can estimate and track the pitch of a speaker using quantitative methods at fine temporal granularities. Pitch, and other low-level descriptors (LLDs) of human behaviors (Schuller et al., 2007), can be extracted using well-developed signal processing methods, which in turn can be mapped to relevant high-level human behavior via machine learning algorithms.

Computer technology has the advantage of automatically analyzing data in a consistent, repeatable manner. In addition, computational algorithms can be incrementally improved, benefiting from more data and improved methodologies. Another obvious advantage of computer technology is that it will not fatigue. Finally, whereas current human behavioral methods are not scalable to coding large amounts of data over long periods of time, computer technology is highly scalable. Technology can also be modularized, with separate algorithms specializing in modeling specific human behaviors, which could make the technology adaptable from one domain of research to another distinct but overlapping domain.

Our aim in this work is to augment the observational power of the researcher and therapist with novel computational tools and techniques. Specifically, we explore the power of objective signal-based measures (speech-derived audio cues), extracted during real marital discussions, in predicting perceptual observations made by evaluators trained on a manual human behavioral coding system. Thus, our goal is to emulate *human* evaluators observing *human* behavior.

This research is part of a growing field, behavioral signal processing (BSP), aimed at better connecting the behavioral sciences with signal processing methods. Traditional signal processing research (e.g., speech recognition, face/hand tracking) concentrated on modeling more objective human behaviors (e.g., "what was spoken?"). BSP builds upon traditional engineering tools and methods to model more abstract human behaviors in realistic scenarios that are especially relevant in psychology and related fields (e.g., the question "is one spouse blaming the other?" in marital therapy).

Significant work related to BSP has concentrated on extracting human-centered information from audio-video signals, including social cues (Vinciarelli et al., 2009), affect and emotions (Lee and Narayanan, 2005; Grimm et al., 2007; Schuller et al., 2009a; Yildirim et al., 2010), and intent (Jurafsky et al., 2009). The increased push to analyze realistic human interactions and naturalistic data (as opposed to acted or artificially constrained data) is most evident in the affective computing and emotion recognition communities (Campbell, 2000; Douglas-Cowie et al., 2003, 2007; Devillers et al., 2005; Devillers and Campbell, 2011; Burkhardt et al., 2009).

This paper builds upon some of our recent work in applying the basic ideas of BSP using the Couple Therapy corpus (Christensen et al., 2004), discussed in detail in Section 2. This corpus consists of recordings of a husband