EJCP The European Journal of Counselling Psychology
ejcop.psychopen.eu | 2195-7614

The Language of Interpersonal Interaction: An Interdisciplinary Approach to Assessing and Processing Vocal and Speech Data

Sarah Weusthoff[a], Garren Gaut[b], Mark Steyvers[b], David C. Atkins[c], Kurt Hahlweg[a], Jasara Hogan[d], Tanja Zimmermann[e], Melanie S. Fischer[f], Donald H. Baucom[f], Panayiotis Georgiou[g], Shrikanth Narayanan[g], Brian R. Baucom*[d]

[a] TU Braunschweig, Braunschweig, Germany. **[b]** University of California, Irvine, CA, USA. **[c]** University of Washington, Seattle, WA, USA. **[d]** University of Utah, Salt Lake City, UT, USA. **[e]** Hannover Medical School, Hannover, Germany. **[f]** University of North Carolina, Chapel Hill, NC, USA. **[g]** University of Southern California, Los Angeles, CA, USA.

Abstract

Verbal and non-verbal information is central to social interaction between humans and has been studied intensively in psychology. Especially, dyadic interactions (e.g. between romantic partners or between psychotherapist and patient) are relevant for a number of psychological research areas. However, psychological methods applied so far have not been able to handle the vast amount of data resulting from human interactions, impeding scientific discovery and progress. This paper presents an interdisciplinary approach using technology from engineering and computer science to work with continuous data from human communication and interaction on the verbal (e.g. use of words, content) and non-verbal (e.g. vocal features of the human voice) level. Text-mining techniques such as topic models take into account the semantic and syntactic information of written text (such as therapy session transcripts) and its structure and intercorrelations. Speech signal processing focuses on the vocal information in a speaker's voice (e.g. based on audio- or videotaped interactions). For both areas, an introduction defining the respective method and related procedures, and sample applications from psychological publications complementing or generating behavioral codes (e.g. in addition to cardiovascular indices of arousal or as a form to encode empathy) are provided. We close with a summary on the opportunities and challenges of learning and applying tools from the novel approaches described in this manuscript to different areas of psychological research and provide the interested reader with a list of additional readings on the technical aspects of topic modeling and speech signal processing.

Keywords: fundamental frequency, topic models, arousal, behavioral signal processing, dyadic interaction, communication

The European Journal of Counselling Psychology, 2018, Vol. 7(1), 69–85, doi:10.5964/ejcop.v7i1.82

Received: 2015-05-20. Accepted: 2018-07-14. Published (VoR): 2018-09-17.

*Corresponding author at: Department of Psychology, University of Utah, 380 S. 1530 E. BEHS 502, Salt Lake City, Utah 84112, USA. phone: +001 (801) 581-5841, fax: +001 (801) 581-8496. brian.baucom@psych.utah

Interpersonal interactions are challenging to study but are central to our daily lives and health and well-being. Consider the example of psychotherapy. A psychotherapy session is a complex interaction between a therapist and a patient, and when successful, something in the dyadic interaction leads to a reduction in psychiatric symptoms and an increase in the patient's functioning and well-being. Yet, the raw 'data' of psychotherapy is the verbal (i.e. acoustic, semantic) and non-verbal behavior of the two people, and these data streams are continuous throughout the psychotherapy session. The inherent complexity of such data has typically led researchers to use methods that drastically simplify the field's understanding of this complex process (Baucom, 2010).

Verbal and non-verbal data streams are not specific to psychotherapy but are common to most interpersonal interactions, which are a focus of study across a wide array of scholarly disciplines. Moreover, researchers use the tools they know and are often unaware of novel methodologies developed in other disciplines. The lack of interdisciplinary collaboration on methods for studying interpersonal interaction has slowed the pace of scientific discovery. In addition, it likely contributes to the propagation of systematic errors associated with the limitations of current methods. Within psychology, the most common method for studying interpersonal interaction is observational behavior coding. In this method, behavior is quantified by human coders (sometimes called annotators or raters) according to rules developed by investigators to capture the essential aspects of the dyadic interaction. Using this methodology, trained coders watch or listen to a dyad's interacting and score them on dimensions defined by the coding system. However, behavioral coding has a number of shortcomings (Baucom & Iturralde, 2012): (a) it is time consuming, often involving months of training prior to the actual work; (b) it can be error prone as human coders do not always agree with each other on coding decisions; (c) behavioral coding does not 'scale up' to larger samples due to time constraints; (d) behavioral coding systems often do not translate cross-culturally (Zimmermann, Baucom, Irvine, & Heinrichs, 2015), impeding replication and generalization of findings; and (e) the coding systems are simplifications of the true complexity of interactions because raters are limited in the quantity and temporal specificity of factors they can observe and code during a dyad's interactions. What is needed is a new set of methods for studying interpersonal interaction that allow us to move beyond these limitations.

Fortunately, methods and tools to work with complex linguistic data exist within engineering and computer science, falling broadly within the categories of speech signal processing and statistical text-mining and natural language processing (Busso, Lee, & Narayanan, 2009; Gaut, Steyvers, Imel, Atkins, & Smyth, 2017). At a fundamental level, signal processing techniques use computer algorithms to derive "informative" quantities from highly multivariate, continuous streams of input. One of the tremendous advantages of signal processing methods is that they can be used to process vast amounts of information far beyond what an individual rater could garner from watching a dyadic interaction. In much the same way that observational coders are trained to recognize and rate classes of behaviors, speech signal processing and statistical text-mining use algorithms and models to estimate mathematical quantities, called features, that characterize aspects of the original signal, either voice or text. Some features provide a psychologically meaningful measurement of behavior in and of themselves (e.g. vocally encoded emotional arousal measured by the fundamental frequency [f_0] of the speech sound wave) while other features can be combined to recognize social behaviors of interest with statistical techniques. Thus, one application of this methodology has been to replicate human coding, but using only acoustic and text inputs. For example, working with transcripts of drug addiction counseling sessions, these methods have led to automated coding of therapists' behaviors during motivational interviewing, including behaviors such as simple and complex reflections, open and closed questions, and affirmation (Atkins, Steyvers, Imel, & Smyth, 2014; Can, Georgiou, Atkins, & Narayanan, 2012). Similar advances have been realized using acoustic data for automated coding of spousal behavior during discussions of a difficult relationship problem (e.g. Lee et al., 2010, 2014). These methods not only represent significant methodological advancements, but also open up new avenues for theoretical refinement and advancement. These new methodologies overcome some of the inherent limitations in behavioral coding and allow researchers to ask familiar questions in new ways and to ask entirely new questions.

Method

Based on the international and interdisciplinary work and collaboration during the VolkswagenStiftung's summer school "The language of interaction," the goal of this manuscript is to provide a general overview of speech signal processing and statistical text-mining methods that have the potential to exponentially increase our ability to understand and conduct research on important dyadic interactions. The methods are introduced using example applications from two types of interpersonal interaction that are central to psychology research: (a) partners within a committed romantic relationship, and (b) patients and therapists in a psychotherapeutic context. Our goal is to introduce the methodologies; however, the current article is not intended as a tutorial, which would be beyond the scope of a single article. Throughout, we point to additional resources to assist learning more about the methods. The current article is also not intended to present a comprehensive review of existing work using these and other allied methods. We close by proposing a research agenda for further integrating these methods into psychological science and clinical practice.

This manuscript only covers research using the methods mentioned above and meeting the following criteria:

Search Strategy

Databases

PsycInfo, PubMed, IEEE Xplore, International Speech Communication Association Online Archive

Search Terms

Keywords

set 1: fundamental frequency couples, set 2: topic models couples, set 3: topic models psychotherapy, set 4: behavioral signal processing.

Reasons behind the keyword selection — These sets of keywords were used to identify manuscripts that used the specific computational methodologies that are the focus of this manuscript.

Criteria for selecting studies — Studies included in this manuscript were selected on the basis of the clarity with which they illustrated one or more methodological techniques within the content of research on romantic relationships, Motivational Interviewing and/or psychotherapy.

Reasons behind the criteria — We established these criteria to be the focus of the manuscript on providing an introduction to what we consider to be the most promising computational methods for working with speech and text data collected in dyadic interaction contexts with strong clinical relevance for researchers and practitioners.

Review and Discussion

Statistical Text-Mining With Topic Models

Traditionally, studying the content of psychotherapy sessions has involved the use of human judgment. One such approach is content analysis (Krippendorff, 2013) wherein the researcher develops a set of categories

with explicit definitions for each category, and human raters manually code the text according to these categories. These inductively derived categories can inform treatment process (i.e. what actually occurs during therapy) and serve as predictors of treatment outcomes (Svartvatten, Segerlund, Dennhag, Andersson, & Carlbring, 2015). However, the significant disadvantage of content analysis, and any method relying on human judgment, is that it is difficult to scale up to large collections of transcripts. One popular method to automatically analyze text is Linguistic Inquiry Word Count (LIWC; Pennebaker, Booth, & Francis, 2007), a program that recognizes 2,300 words and classifies them within 70 predefined classes (e.g. negative and positive emotion words). LIWC has been used to study emotional, cognitive, structural, and process components present in verbal and written speech (Tausczik & Pennebaker, 2010). At the same time, LIWC ignores the context in which the words occur, and because the words and categories are fixed, there is no ability to adapt to a particular domain.

A more recent approach to analyzing the text in psychotherapy transcripts is based on topic models (Atkins et al., 2012; Atkins et al., 2014; Imel, Steyvers, & Atkins, 2015). Topic models (also called latent Dirichlet allocation) are a type of latent mixture model, in which the model identifies groups of words that reliably co-occur across therapy sessions (or "documents", more generally). The model represents the observed transcripts as mixtures of various underlying topics. For example, a transcript from a drug intervention might be represented as a mixture of a drug use topic and a rehabilitation topic where words such as drug or problems are highly likely to be expressed in the drug use topic, while words such as healing and change are more likely to be expressed in the rehabilitation topic. The resulting topics can be used to identify themes in text and also as data reduction of text for predictive modeling. Topic models share similarities with other dimensionality reduction techniques such as Latent Semantic Analysis (Landauer & Dumais, 1997), cluster analysis, or principle component analysis but were specifically designed to create interpretable dimensions when applied to text. This is advantageous when the goals are not just dimensionality reduction or prediction but also interpretation of the underlying patterns of the linguistic data.

Topic models applied to psychotherapy transcripts have sometimes been used for exploratory data analysis, where the goal is to summarize, explore, and discover the types of topics that are discussed (called unsupervised learning in the machine learning literature; Hastie, Tibshirani, & Friedman, 2009). Alternatively, they can also be used to predict some variable of interest based on the text in the transcript, such as behavioral codes or treatment outcomes (called supervised learning in the machine learning literature).

Topic Models for Exploratory Data Analysis

In the unsupervised learning case, the model associates individual words with latent topics by analyzing the statistical co-occurrences between words across transcripts. Words that tend to co-occur with other words tend to be placed in the same topic. For example, Table 1 shows topics inferred from a large randomized clinical trial of two behaviorally-based couple therapies (Atkins et al., 2012). The table shows a few illustrative topics the model inferred and the 15 words that were most frequently assigned to each topic. Like factor analysis, the topic names (in bold typeface) are labels supplied by the authors and are not automatically learned by the model.

Table 1

Examples of Topics From a Topic Model Applied to Couple-Therapy Transcripts and the 15 Most Likely Words for Each Topic. Adapted From Atkins et al. (2012)

Family	Finances	Relationships	Sex	Transportation	Work
mom	money	married	sex	car	job
mother	dollars	together	sexual	drive	work
dad	buy	relationship	love	park	career
sister	hundred	date	part	down	money
brother	card	live	interesting	street	day
call	thousand	met	initially	turn	company
day	bought	move	touch	traffic	support
father	credit	attracted	back	accident	people
deal	pay	remember	desire	home	situation
live	fifty	marriage	physical	bus	hours
parents	cost	thought	intimacy	drove	week
down	give	months	life	hours	positive
care	car	two	talk	direction	happy
stuff	expensive	decided	relationship	freeway	business
house	five	pretty	pleasure	walk	interview
Negative Emotional Content			**Positive Emotional Content**		
angry	give	upset	good	[laugh]	good
anger	shit	back	thought	guess	nice
hurt	pissed	mad	[laughing]	good	thought
frustrated	point	temper	pretty	work	felt
trying	whatever	talk	people	thank	appreciate
upset	man	crying	talk	give	week
mad	fuck	angry	part	wow	remember
point	care	sorry	summer	definitely	day
sad	god	understand	enjoy	[laughing]	couple
emotional	black	fact	remember	back	notice
part	problem	late	fun	obviously	work
felt	fine	apologize	u	you'll	great
whatever	walk	fine	nice	hard	pretty
express	white	ready	great	relax	thank
respond	cannot	reason	vacation	[all laugh]	realize

Note. The bold-face labels are subjective interpretations of the word clusters and are not automatically determined.

The topic model also produces a distribution of topics over sessions (or whatever defines the basic textual unit of analysis). With these probability distributions, it is possible to investigate temporal changes in topics over sessions or identify individual talk turns with specific content (e.g. particular interventions or important topics such as drug and alcohol use).

Imel et al. (2014) used topic models to compare the linguistic similarity in psychotherapy sessions (N = 1,318), comparing four different types of treatment (Medication Management, Psychodynamic therapy, Cognitive-Behavioral therapy, Humanistic/Existential therapy). An unsupervised topic model was used, and the resulting topics were used to quantitatively summarize the semantic content of each session, which in turn was used to compare each session with every other session. Figure 1 shows a multi-dimensional scaling of the topic model

derived semantic summaries where each point represents a single session. Sessions conducted using the same treatment type tend to be linguistically similar, although, there was some variability within each treatment type. The medication management session outlined by a black circle was more similar to Humanistic/Experimental therapy or CBT than other medication management trials. Inspection of the transcript revealed that there was no direct discussion of medication or dosage and that the session was more focused on providing psychotherapy. This demonstration shows how topic models can be used to compare psychotherapy session across treatment type.

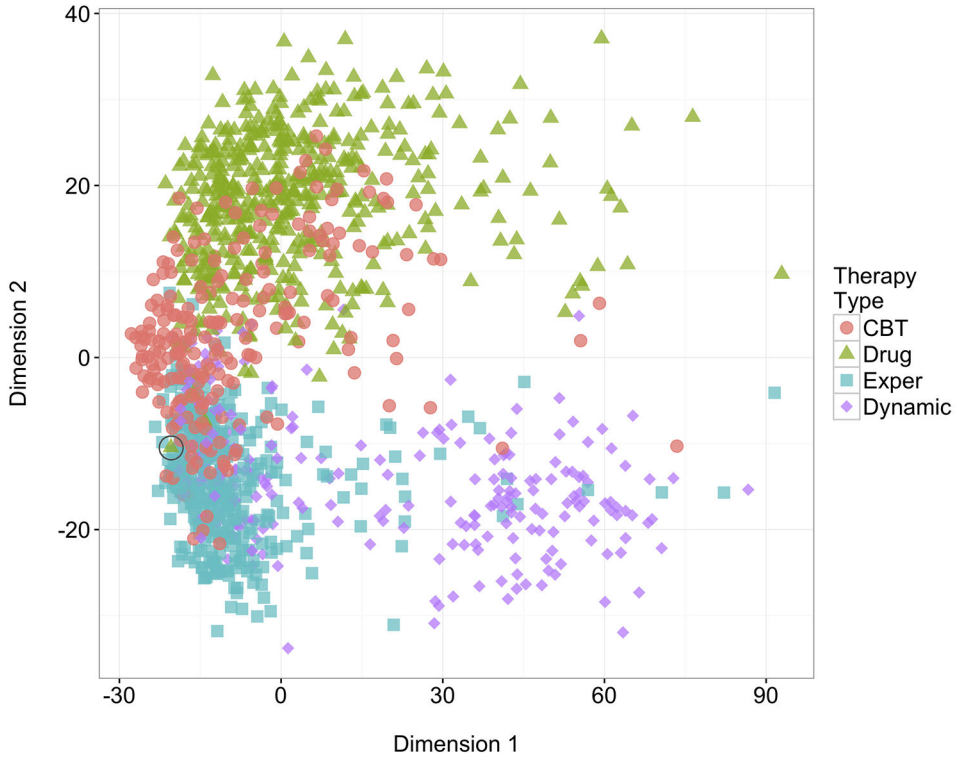

Figure 1. Multidimensional scaling of 1318 sessions in a 200 topic space. Colors correspond to different treatment approaches. One outlier medication management session is circled in black. Adapted from Imel et al. (2014).

Topic Models for Behavioral Coding

An extension of the basic, unsupervised topic model is called a labeled topic model and can be used to predict behavioral codes or content discussed in treatment (Atkins et al., 2014; Gaut et al., 2017). The topics in the model are placed in correspondence with behavioral or content codes (e.g. subject or symptom) or latent background codes that do not correspond to any of the observable codes. The model learns which words are associated with each topic and can infer which content or behavioral codes are representative for individual utterances, talk turns, or sessions.

For example, Table 2 shows output from a labeled topic model applied to the same psychotherapy corpus used by Imel et al. (2014). However, for this analysis, 209 content or symptom codes were included in the model fitting process (Gaut et al., 2017). The model learned words that are indicative of each code and even found unexpected semantic relationships between words and codes, such as grouping the words ‘dishes’ or ‘cats’ in the topic corresponding to a code about irritability. This model included latent background topics to capture lin-

guistic variation not specific to any one code, and these topics correspond to broad concepts such as family, work, home, fitness, and sleeping routine. Gaut et al. (2017) also demonstrated that the labeled topic model can find local information within a session that is associated with a session-level code or tag. For example, from a session tagged with a 'suicidal behavior' code, the model was able to disentangle talk turns that were specifically associated with suicide from talk turns associated with other unrelated topics. Comparing the labeled topic model to human-produced codes using area under the ROC curve (AUC), the labeled topic model predicted subject and symptom codes, showing strong concordance with human-produced codes.

Table 2

Example Topics Learned Using Supervised Topic Models

Type of Code / Code	Most Likely Words from Inferred Topic Distribution
	Gaut et al., submitted
Subject	
Medications	medicine, mg, dose, wellbutrin, medicines, lamictal, prescription, effects, side_effects, ability
Relationships	relationship, women, feels, friend, relationships, boyfriend, date, position, example, react
Parent-child relations	mother, father, love, remember, relationship, parents, brother, emotional, loved, needed
Depressive disorder	depression, medication, doctor, medicine, prozac, depressed, zoloft, generic, wellbutrin, add
Spousal relationships	wife, marriage, married, husband, relationship, mhm, children, attitude, divorce, got_married
Symptom	
Anxiety	anxiety, anxious, panic, nervous, depression, worried, worst, fine, experience, helps
Depression	depressed, depression, doctor, pain, die, needed, drugs, low, xanax, mg
Anger	angry, feelings, anger, express, get_angry, be_angry, reaction, feels, pissed, 'm_feeling
Low self-esteem	love, teaching, boyfriend, positive, stupid, attractive, fit, negative, sorta, criticism
Irritability	annoyed, irritable, message, safe, dishes, cause, wife, skin, irritated, cats
Background	
background 9	friends, family, mom, dad, close, sister, brother, daughter, men, lives
background 13	care, stop, took, weight, takes, ready, lose, take_care, amount, body
background 23	house, room, walk, bed, door, walking, rid, front, throw, clean
background 36	job, wants, work, business, works, office, busy, baby, buy, paper
background 39	morning, sleep, hours, friday, sleeping, monday, tomorrow, saturday, wake, bed
	Atkins et al. 2014
MISC Code	
Question Closed	have you, questions, would you, are you, risk, drink, you think, okay so, do you have, heard
Question Open	what do you, what do, group, do you think, you think, expected, how do you, bar, how do
Reflection Simple	sounds, it sounds like, mentioned, okay so, sounds like, it sounds, like you, you said
Reflection Complex	sounds, it sounds like, sounds like, it sounds, sounds like you, like you, so it sounds
Study Type	
ARC	alcohol, drinking, drink, calories, positive, negative, drinks, level, a lot of, you're drinking
ESPSB	break, spring, drinks, spring break, sex, standard, drink, student, average, number
ESP21	birthday, drinks, wine, drink, standard, beer, alcohol, your birthday, twenty first birthday
HMCBI	drug, you know, years, pain, using, drugs, you know what, know what i, marijuana, you know i
iCHAMP	marijuana, smoke, use, smoking, sleep, rem, smoked, marijuana use, evergreen, month

Note. ARC = alcohol intervention study; ESPSB = alcohol intervention for spring break study; ESP21 = alcohol intervention for students turning 21 study; HMCBI = drug intervention study; iCHAMP = marijuana intervention study.

The previous study examined generic codes, focused on common discussion topics or patient symptoms that might occur in psychotherapy sessions. Atkins et al. (2014) examined the utility of a labeled topic model to generate a much more specific type of observational code, fidelity codes for motivational interviewing (MI). They trained the model using 148 sessions from five MI intervention studies that were transcribed and coded using the Motivational Interviewing Skills Code (MISC; Miller, Moyers, Ernst, & Amrhein, 2008). The studies were heterogeneous in focus including alcohol, marijuana, and poly-drug targets. The model learned topics for 12 MISC behavioral codes, as well as for each of the five different studies (see Table 1 for examples). The topics for MISC behavioral codes and specific studies generally have face-valid interpretations. For example, topics corresponding to asking questions contain phrases such as ‘have you’, ‘questions’, ‘would you’, and ‘are you’ and topics for the ESPSB study that focused on spring break drinking included words related to partying during spring break such as ‘break’, ‘spring’, ‘drinks’, and ‘spring break’. Table 3 shows the most probable talk turns for a sample of MISC behavioral codes.

Table 3

Example of Talk Turns Assigned by the Model for a Sample of Behavioral Codes in Atkins et al. (2012)

Study / Topic	Example Therapist Talk Turns Assigned by Model
Atkins et al. (2012)	
Closed Question	Yeah does that surprise you. Does that sound about right. So well did you have any other questions for me or.
Open Question	So what do you make of that. What do you mean by that. What do you mean by absurd.
Simple Reflection	Okay and you said that you used to drink a little bit more last year it sounds like. Yeah you mentioned that and you felt like that kinda. It sounds like it it sounds like you're doing it it sounds like you have um you say you feel better right is that physically and emotionally.
Complex Reflection	Mm-hmm so you're two birds of a feather it sounds like. Yeah it sounds like you've turned it around.

Atkins et al. (2014) tested the ability of the model to predict behavioral codes by computing scores for each talk turn in each session that corresponded to MISC behavioral items. They assessed the ability of the model to predict codes assigned by raters and compared model performance to human reliability. To determine predictive ability, Atkins et al. (2014) computed the area under the curve (AUC) for each code. The model coded talk turns with an average AUC of 0.73 and tended to perform best at coding open and closed questions, complex reflections, affirmations, structure, and empathy. The model performed significantly better than chance guessing and generally performed better when codes are tallied across sessions as opposed to at each individual talk turn. On several codes (e.g. complex reflections, information giving), the model reliability was comparable to human reliability but for other codes, human reliability was significantly better than model performance. This suggests that the labeled topic model can be competitive with human raters for some codes, but that other codes may be more difficult to capture in a topic modeling framework.

Future Directions in Text Analysis of Dyadic Interaction

Topic models primarily focus on word co-occurrence within documents (e.g. sessions, talk turns) and ignore any temporal dependence between words. Statistical text-mining for psychotherapy transcripts could be improved by incorporating information about temporal and syntactic structure. Recent neural network methods attempt to model this dependence by modeling words as high-dimensional vectors whose representation depends on surrounding context words (Mikolov, Corrado, Chen, & Dean, 2013), as well as syntactic information (Socher, Bauer, Manning, & Ng, 2013). Other methods model words as high-dimensional probability distributions where semantic meaning can be captured by set relations with other words (Vilnis & McCallum, 2015). These methods have shown promise in several natural language processing tasks such as sentiment analysis but have yet to be applied to psychotherapy transcripts.

Speech Signal Processing

Whereas, statistical text-mining primarily addresses “what was said” in a conversation, psychologists have commonly used observational coding to also measure “how it was said.” The application of observational coding systems needs to take into account a number of different aspects of the interaction, such as the communication setting, target population, or specific behaviors displayed by the interaction partners on different levels (i.e. micro- versus macro-analytic coding; for further details see Humbad, Donnellan, Klump, & Burt, 2011), adding to the complex nature of the observational coding practice itself. However, what unites coding systems focusing on non-verbal, vocal aspects of interpersonal interactions is the attempt to capture the paralinguistic components of spoken language by taking into account vocal aspects such as voice tone, loudness, or intonation.

Other disciplines have been developing computational methods for quantifying information communicated in the voice, but these methods have only recently been applied in psychological research. The emerging field of behavioral signal processing (BSP; Narayanan & Georgiou, 2013) is developing methods to decipher the complex and heterogeneous chain of events that comprises human behavior and transform it into its basic signal components, making it analyzable using tools from engineering and computer science. Speech signals expressed in the human voice are especially informative as they convey both the content of a message (verbal/linguistic information) and paralinguistic information about the psychological state of the speaker using non-verbal or vocal cues (e.g. prosody; Juslin & Scherer, 2005).

In order to communicate with another human being, a speaker needs to generate a message in a form that can be understood by others. In the case of speech production, the neuro-muscular controls start and coordinate the interplay of respiratory, phonetic, and articulatory muscles in the vocal tract system to provide the anatomical and physiological sources for generating a continuous acoustic speech sound wave with the desired and / or necessary characteristics for the environment the individual is communicating in (e.g. appropriate loudness). The recipient of such a message processes the acoustic waveform of the speech signal via the basilar membrane and resulting neural transduction into its discrete features, which are then coded into phonemes, words, and sentences that can be interpreted and understood by the listener (Narayanan, 2014).

For speech signal processing purposes, the discrete features of the continuous speech sound wave comprising the physical basis of any spoken message are analyzed. To be able to extract this information, the speech sound wave is periodically and repeatedly segmented and analyzed with regard to a feature of interest, yielding

a numerical index for each segment of analysis. One such example could be vocal fundamental frequency, or f_0, scores (perceived as the voice pitch of a speaker) for any given time period of interest. A number of other vocal features based on different aspects of the human voice can be analyzed as well (Juslin & Scherer, 2005, for a review). Research in dyadic interactions, however, has focused on f_0 as one of the most emotionally salient aspects of human voice (Busso et al., 2009), and, thus, presents the scope of this manuscript.

Behavioral Signal Processing of Vocally Encoded Emotional Arousal

One potential application of speech signal processing results from the Component Process Model of Emotion (CPME; Scherer, 2009). The CPME views emotions as recurring, prototypical, and adaptive reactions for goal-achieving behavior (Buss, 2005) that influence various bodily systems physiologically, among them the human voice (Juslin & Scherer, 2005) and its acoustic features and related signals, such as f_0.

According to the CPME, an individual evaluates an internal or external event and its consequences on different levels by the interaction of cognitive functions (termed cognitive appraisal). Cognitive appraisal reflects the individual's subjective perception of a situation, object, or event. If the event is appraised as being meaningful for an individual's goals and / or motives, emotional processes (including motivational changes) occur and lead to physiological changes in the autonomic (e.g. cardiac or respiratory system) and somatic (motor-driven expressions in face, voice, or body posture) nervous system. The pattern of activation in these systems that arises during an emotional episode is influenced by numerous factors including the valence (i.e., degree of pleasantness vs. unpleasantness) and arousal (i.e., degree of activation vs. deactivation) of the emotion.

Emotional arousal is expressed through a number of channels such as facial expressions, word usage, and prosodic features of the voice (e.g. fundamental frequency; for a more detailed description and empirical findings concerning the CPME; please see Scherer, 2004, 2005). Vocal fundamental frequency is an acoustic property of the human speech sound wave that can be assessed mechanically and, thus, analyzed and interpreted objectively. During phonation, the first phase of human speech production, air is released from the lungs. The outward flow of air passes over the vocal folds in the larynx, which can be positioned and flexed by muscles under voluntary control. The different levels of tension in the vocal folds lead to vibrations in the passing air (Juslin & Scherer, 2005). The lowest harmonic frequency of these patterns is called (vocal) fundamental frequency (or f_0), measured in cycles per second (Hertz, Hz) over the time period of interest. Higher tension in the vocal folds corresponds to higher vibration rates and to higher f_0 values (Weusthoff, Baucom, & Hahlweg, 2013). On the auditory level, perceived voice pitch is analogous to fundamental frequency, and higher f_0 is perceived as a higher voice pitch (Frick, 1985).

Emotions and emotional arousal displayed in the human voice (Busso, Lee, & Narayanan, 2009) seem to be an especially flexible tool in passing on information about an individual's result of appraisal (e.g. in the case of sensing danger in one's environment; Juslin & Laukka, 2003). Since the human voice is one of the sounds most often experienced by human beings in life (Belin, Zatorre, & Ahad, 2002), it is very useful in studying human interactions per se, and especially important in psychological research as it provides a continuous stream of information that is interpreted and acted upon by both interaction partners.

Weusthoff, Baucom, and Hahlweg (2013) aimed to clarify what forms of information are encoded in f_0 during naturalistic dyadic interactions and to investigate whether f_0 provides a valid form of assessment of emotional arousal. F_0 range is a specific index of f_0 calculated as the difference between an individual's maximum and

minimum f_0 score during a time period of interest. During couple conflict discussions, f_0 range and its concurrent links to other well-established indices of emotional arousal (cardiovascular and endocrine), and other important aspects of couple functioning (self-reported and observed communication behavior) were investigated. The sample consisted of *N* = 67 severely distressed, heterosexual German couples participating in couple-relationship education (namely, *EPL – Ein Partnerschaftliches Lernprogramm*; Hahlweg, Markman, Thurmaier, Engl, & Eckert, 1998).

As expected, positive associations emerged between f_0 and physiological variables (in this study, heart rate, blood pressure, and cortisol): Higher f_0 range was associated with higher levels of all physiological indicators of emotional arousal. Also, communication behavior was significantly related to f_0 range. Namely, higher levels of f_0 range were associated with higher levels of self-reported negative communication behavior. For observed communication behavior, higher levels of f_0 range were linked to higher levels of negative communication behavior and lower levels of positive communication behavior. Additionally, simultaneous examination of physiological variables and observationally-coded communication behaviors revealed that associations between both sets of variables and f_0 range were largely independent of one another. Although male and female f_0 range were significantly different from each other (with females having higher f_0 range values than males), no significant gender differences emerged in any of the predictors of interest associated with f_0 range mentioned above.

This collection of findings suggests that f_0 range may be most reasonably interpreted as a vocal distress signal during couple conflict, independently reflecting both socially learned ways of interacting with others (communication behavior) and basic processes in physiological responding (heart rate, blood pressure, cortisol). The findings demonstrate that these associations reflect variability in response due to a particular conflict, a general style of responding during conflict, and individual differences.

Behavioral Signal Processing for Behavioral Coding

Another application of speech signal processing is the use of BSP methods to generate observational coding data. Using BSP methods to generate observational coding data from acoustic features is similar in many ways to the labeled topic models described above. In both cases, a large number of input features (e.g. words, f_0) are used to predict what score an observational coder would give to an interaction using a supervised learning model. The primary distinction we are making here is that it is possible to use linguistic features, acoustic features, or a combination of the two for this purpose.

One way that acoustic and linguistic features differ is in the methods used to extract features. Acoustic features are generally derived from continuous waveforms and the methods used to extract features from this type of data, therefore, reflect the unique properties of continuous waves. This process is analogous to how heart rate (a feature) is extracted from an ECG recording (a continuous signal). Heart rate is typically measured as the number of R-spikes that occur during a 60 second window where an R-spike is defined as a most prominent, upward inflection that occurs between two smaller upward inflections and a downward inflection. In much the same way, there are numerous acoustic features that can be used to index spectral, prosodic, and vocal quality aspects of recorded speech.

BSP for observational coding works by extracting a large number of acoustic features and “learning” in what way those features are related to observational coding data. For example, researchers have shown that it is possible to predict codes from both the Couple Interaction Rating Scale 2 (CIRS-2; Heavey, Gill, & Christensen,

2002) and the Social Support Interaction Rating System (SSIRS; Jones & Christensen, 1998) using BSP methods with acoustic features (Black, Georgiou, Katsamanis, Baucom, & Narayanan, 2011). CIRS2 and SSIRS codes are used to quantify spouses' behaviors during problem-solving discussions and include ratings of positive and negative behaviors and affect. A comparison of the resulting BSP-generated codes and human-generated codes showed an average accuracy of 75% for rating wives' behaviors and 73% for rating husbands' behaviors, with even higher accuracy for specific codes. Negativity in husbands was rated with 86% accuracy, for example. Advances continue to be made in the use of BSP for such complex codes, and accuracy will only improve as more research is done in this area.

Future Directions

We have provided an application-focused overview of statistical text-mining and speech signal processing, highlighting how they can be used to tap similar types of information as behavioral codes and even be used as methodologies to generate behavioral codes. In this final section, we discuss key considerations in using these methodologies in psychological science, as well as future directions broadly.

No methodology is ideal for all applications, and there must be thoughtful, theory-based applications of these methodologies within psychological science. For example, Imel et al. (2014) used f_0 to study the process of empathy in motivational interviewing sessions and used a coherent theoretical understanding of empathy to do so. In particular, the perception-action model of empathy (Preston & De Waal, 2002) has emphasized physiological synchrony as a core component of empathy. Thus, Imel et al. (2014) used f_0 as a marker of vocally encoded arousal and demonstrated that it is more highly correlated (across one-minute intervals and entire sessions) within MI sessions rated as highly empathic versus those that were rated low on empathy. Baucom, Atkins, and Christensen (2010, as cited in Baucom & Atkins, 2012) also examined covariation in f_0 to examine a theoretically derived, interpersonal emotional process, in this case interpersonal emotion regulation. Baucom et al. (2010) examined covariation in spouse's f_0 while they were discussing an area of disagreement in their relationship. Polarization theory (e.g. Baucom & Atkins, 2012) suggests that a key characteristic of relationship distress is difficulty regulating emotion during relationship conflict. Consistent with this idea, Baucom et al. (2010) found that stronger cross-partner associations in f_0 were associated with poorer interpersonal emotion regulation. Considering these examples jointly demonstrates that while there are significant cross-person associations in f_0 during many, if not most, interactions, these associations do not represent the same process. This example illustrates that these methodologies can be powerful new tools for studying human behavior and dyadic interaction, but as always, they must be used within the context of psychological theory.

Beyond theoretical consideration, who can use these methods? Is it reasonable to think that psychologists could effectively learn these methodologies on their own, or do they require collaborating with colleagues from engineering and computer science? Our informed opinion is that both are possible, though the latter is preferable. Consider statistics as an analogy. All psychologists (and most other social scientists) learn statistical methods in their training and have varying degrees of facility and expertise in applying data analytic techniques. At the same time, it is very challenging for psychologists to be expert in the many areas of statistical methodology and stay current on new developments. This is largely true for speech signal processing and statistical text-mining. Psychologists with interests and skills in quantitative methodologies should certainly be encouraged to learn how to estimate and analyze f_0 or to apply basic topic models to their text data. At the same time, there are significant advantages to collaborating with colleagues who are experts in these areas, as the underlying

math and algorithms (e.g. fast Fourier transform for spectral features from voice, Dirichlet distribution as key component of topic models) are not common to most psychology training.

Summary

More than one hundred years ago, John Watson made a powerful argument for the value of new methods for studying behavior in noting that:

> As our methods become better developed it will be possible to undertake investigations of more and more complex forms of behavior. Problems which are now laid aside will again become imperative, but they can be viewed as they arise from a new angle and in more concrete settings. (Watson, 1913, p. 175)

The most exciting aspect of the methods introduced in this manuscript is their potential to fulfill Watson's vision. Continued development and advancement of these methods will be challenging and necessarily involve tight interdisciplinary collaboration. However, such advancement has tremendous potential to make break-through discoveries because it would enable the field to address previously intractable problems and to do so in rich and thoughtful ways. For example, these computational methods have the potential to be used to monitor adherence to psychotherapy protocols in healthcare networks (e.g. Imel, Steyvers, & Atkins, 2015) and to provide private practice clinicians with a means for objectively assessing treatment progress and outcomes (Baucom & Iturralde, 2012).

One important element of interdisciplinary collaboration on these methods is sufficient understanding to the mechanics involved in the methods themselves. As noted above, it is not necessary for psychologists to be able to develop the computer code or mathematical algorithms underlying these methods, but knowledge of how the methods work increases a psychologist's ability to collaborate effectively. With this idea in mind, we close with a list of references for further technical readings related to the computational methods that were introduced in this manuscript.

Funding

This research was supported by grants from the VolkswagenStiftung awarded to Kurt Hahlweg, David C. Atkins, and Brian R. Baucom (Az.: 88 374), and by the Land Niedersachsen to Kurt Hahlweg (Niedersachsenprofessur 65+, ZN 2795, 15-2/12).

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

References

Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R., & Christensen, A. (2012). Topic models: A novel method for modeling couple and family text data. *Journal of Family Psychology, 26*, 816-827. doi:10.1037/a0029607

Atkins, D. C., Steyvers, M., Imel, Z. E., & Smyth, P. (2014). Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science, 9*, Article 49. doi:10.1186/1748-5908-9-49

Baucom, B. R. (2010). Power and arousal: New methods for assessing couples. In K. Hahlweg, M. Grawe-Gerber, & D. H. Baucom (Eds.), *Enhancing couples: The shape of couple therapy to come* (pp. 171-184). Cambridge, MA, USA: Hogrefe.

Baucom, B. R., & Atkins, D. C. (2012). Polarization in marriage. In M. Fine & F. Fincham (Eds.), *Family theories: A content-based approach* (pp. 145-166). New York, NY, USA: Routledge.

Baucom, B. R., Atkins, D. C., & Christensen, A. (2010). *Changes in vocally encoded emotional arousal in two behavioral couple therapies*. Paper presented at the triannual meeting of the World Congress of Behavioral and Cognitive Therapies, Boston, MA, USA.

Baucom, B. R., & Iturralde, E. (2012, January). *A behaviorist manifesto for the 21st Century.* Paper presented at the Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), Hollywood, CA, USA.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*, 17-26. doi:10.1016/S0926-6410(01)00084-2

Black, M. P., Georgiou, P. G., Katsamanis, A., Baucom, B. R., & Narayanan, S. S. (2011). “You made me do it”: Classification of blame in married couples’ interactions by fusing automatically derived speech and language information. In *Proceedings of Interspeech, Florence, Italy* (pp. 89-92). Retrieved from https://sail.usc.edu/publications/files/is2011_black_coupther_paper.pdf

Buss, D. M. (2005). *The handbook of evolutionary psychology.* Hoboken, NJ, USA: Wiley.

Busso, C., Lee, S., & Narayanan, S. (2009). Analysis of emotionally salient aspects of fundamental frequency for emotion detection. *IEEE Transactions on Audio, Speech, and Language Processing, 17*, 582-596. doi:10.1109/TASL.2008.2009578

Can, D., Georgiou, P. G., Atkins, D. C., & Narayanan, S. S. (2012, September). *A case study: Detecting counselor reflection in psychotherapy for addictions using linguistic features.* Paper presented at the INTERSPEECH 2012: 13th Annual Conference of the International Speech Communication Association, Portland, OR, USA.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*, 412-429. doi:10.1037/0033-2909.97.3.412

Gaut, G., Steyvers, M., Imel, Z. E., Atkins, D. C., & Smyth, P. (2017). Content coding of psychotherapy transcripts using labeled topic models. *IEEE Journal of Biomedical and Health Informatics, 21*, 476-487. doi:10.1109/JBHI.2015.2503985

Hahlweg, K., Markman, H. J., Thurmaier, F., Engl, J., & Eckert, V. (1998). Prevention of marital distress: Results of a German prospective longitudinal study. *Journal of Family Psychology, 12*, 543-556. doi:10.1037/0893-3200.12.4.543

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY, USA: Springer. doi:10.1007/978-0-387-84858-7

Heavey, C., Gill, D., & Christensen, A. (2002). *Couples Interaction Rating System – 2nd Edition (CIRS-2).* Unpublished manuscript, University of California, Los Angeles, CA, USA.

Humbad, M. N., Donnellan, M. B., Klump, K. L., & Burt, S. A. (2011). Development of the brief romantic relationship interaction coding scheme (BRRICS). *Journal of Family Psychology, 25*, 759-769. doi:10.1037/a0025216

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology, 61*, 146-153. doi:10.1037/a0034943

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy: Scaling up the evaluation of patient provider interactions. *Psychotherapy, 52*, 19-30. doi:10.1037/a0036841

Jones, J., & Christensen, A. (1998). *Couples interaction study: Social support interaction rating system.* Unpublished manuscript, University of California, Los Angeles, CA, USA.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Scherer, K. (2005). Vocal expression of affect. In J. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavioral research* (pp. 65-136). New York, NY, USA: Oxford University Press.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology*. London, United Kingdom: Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240. doi:10.1037/0033-295X.104.2.211

Lee, C. C., Black, M., Katsamanis, A., Lammert, A., Baucom, B., & Christensen, A., … Narayanan, S. (2010, September). *Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples*. Paper presented at the INTERSPEECH 2010: 11th Annual Conference of the International Speech Communication Association, Chiba, Japan.

Lee, C. C., Katsamanis, A., Black, M. P., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2014). Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions. *Computer Speech & Language, 28*, 518-539. doi:10.1016/j.csl.2012.06.006

Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). *Efficient estimation of word representations in vector space.* Retrieved August 3, 2018 from arXiv.org e-Print archive: https://arxiv.org/abs/1301.3781

Miller, W. R., Moyers, T. B., Ernst, D. B., & Amrhein, P. C. (2008). *Manual for the Motivational Interviewing Skill Code (MISC), Version 2.1*. Albuquerque, NM, USA: The University of New Mexico.

Narayanan, S. (2014). *Fundamentals of speech signal processing* [Powerpoint slides]. Retrieved from https://docs.google.com/file/d/0B3gXh02wTJu4UmlrcGhKRERXMlE/edit?pli=1

Narayanan, S., & Georgiou, P. G. (2013). Behavioral signal processing: Deriving human behavioral informatics from speech and language. *Proceedings of the IEEE, 101*, 1203-1233. doi:10.1109/JPROC.2012.2236291

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC* 2007. Austin, TX, USA: Pennebaker Conglomerates.

Preston, S. D., & De Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1-20. doi:10.1017/S0140525X02000018

Scherer, K. R. (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. In A. S. R. Manstead, N. H. Frijda, & A. H. Fischer (Eds.), *Feelings and emotions: The Amsterdam symposium* (pp. 136-157). Cambridge, United Kingdom: Cambridge University Press.

Scherer, K. R. (2005). Unconscious processes in emotion: The bulk of the iceberg. In P. Niedenthal, L. Feldman-Barrett, & P. Winkielman (Eds.), *The unconscious in emotion* (pp. 312-334). New York, NY, USA: Guilford Press.

Scherer, K. R. (2009). Emotions are emergent processes: They require a dynamic computational architecture. *Philosophical Transactions of The Royal Society: B. Biological Sciences, 364*, 3459-3474. doi:10.1098/rstb.2009.0141

Socher, R., Bauer, J., Manning, C. D., & Ng, A. Y. (2013). Parsing with compositional vector grammars. In *51st Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference* (Vol. 1: Long Papers, pp. 455-465). Madison, WI, USA: Omnipress.

Svartvatten, N., Segerlund, M., Dennhag, I., Andersson, G., & Carlbring, P. (2015). A content analysis of client e-mails in guided internet-based cognitive behavior therapy for depression. *Internet Interventions, 2*, 121-127. doi:10.1016/j.invent.2015.02.004

Tausczik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24-54. doi:10.1177/0261927X09351676

Vilnis, L., & McCallum, A. (2015). *Word representation via Gaussian embedding.* Retrieved August 3, 2018 from arXiv.org e-Print archive: https://arxiv.org/abs/1412.6623

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*, 158-177. doi:10.1037/h0074428

Weusthoff, S., Baucom, B. R., & Hahlweg, K. (2013). Fundamental frequency during couple conflict: An analysis of physiological, behavioral, and sex-linked information encoded in vocal expression. *Journal of Family Psychology, 27*, 212-220. doi:10.1037/a0031887

Zimmermann, T., Baucom, D. H., Irvine, J., & Heinrichs, N. (2015). Cross-country perspectives on social support in couples coping with breast cancer. *Frontiers in Psychological and Behavioral Science, 4*(4), 52-61.

Appendix: Recommendations for Additional Reading

Natural Language Processing

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA, USA: MIT Press.

Topic Models

Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R., & Christensen, A. (2012). Topic models: A novel method for modeling couple and family text data. *Journal of Family Psychology, 26,* 816-827. https://doi.org/10.1037/a0029607

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3,* 993-1022.

Vocal Communication of Affect and Emotion

Schuller, B., & Batliner, A. (2014). *Computational paralinguistics: Emotion, affect and personality in speech and language processing*. Chichester, United Kingdom: Wiley.

Owren, M. J., & Bachorowski, J. A. (2007). Measuring vocal acoustics. In J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 239-266). New York, NY, USA: Oxford University.

Behavioral Signal Processing

Narayanan, S., & Georgiou, P. G. (2013). Behavioral signal processing: Deriving human behavioral informatics from speech and language. *Proceedings of the IEEE, 101,* 1203-1233. https://doi.org/10.1109/JPROC.2012.2236291

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy: Scaling up the evaluation of patient provider interactions. *Psychotherapy, 52*, 19-30. https://doi.org/10.1037/a0036841

Machine Learning

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY, USA: Springer.

PsychOpen GOLD is a publishing service by
Leibniz Institute for Psychology Information (ZPID),
Trier, Germany. www.leibniz-psychology.org