A LANGUAGE-BASED GENERATIVE MODEL FRAMEWORK
FOR BEHAVIORAL ANALYSIS OF COUPLES' THERAPY

Sandeep Nallan Chakravarthula[1], Rahul Gupta[1], Brian Baucom[2], Panayiotis Georgiou[1]

[1]University of Southern California, Los Angeles, CA, USA
[2] The University of Utah, Department of Psychology, UT, USA

ABSTRACT

Observational studies for psychological evaluations rely on careful assessment of multiple behavioral cues. Recent studies have made good progress in automating the psychological evaluation, which often involved tedious manual annotation of a set of behavioral codes. However, the current methods impose strict and often unnatural assumptions for evaluation. In this work, we specifically investigate two goals: (1) Human behavior changes throughout an interaction and better models of this evolution can improve automated behavioral annotation and (2) Human perception of this evolution can be quite complex and non-linear and better techniques than averaging need to be investigated. For this purpose, we propose a Dynamic Behavior Modeling (DBM) scheme, which models a spouse as undergoing changes in behavioral state within a session, and contrast it against a Static Behavior Model (SBM) which allows only a constant session-long behavioral state. We use Negativity in a couples therapy task as our case study. We present results and analysis on both models for capturing the local behavior information and predicting the session level negativity label.

Index Terms— Behavioral Signal Processing, Static Behavioral Models (SBM), Dynamic Behavioral Model (DBM), Hard Expectation Maximization algorithms.

1. INTRODUCTION

Observational studies are the basis of most psychological evaluations and they rely on careful observation and assessment of social, affective, and communicative behavioral cues. In family studies research and practice, our topic of interest, psychologists rely on a variety of established coding standards [1–3], with the aim of producing accurate, consistent ratings of human behavior by human annotators. This manual coding is a costly and time consuming process. First, a detailed coding manual must be created, which often requires several design iterations. Then, multiple coders, each of whom has his/her own biases and limitations, must be trained in a consistent manner. The process is mentally straining and the resulting human agreement is often quite low [4].

Great strides have been made over the last few years on establishing the ability of Signal Processing and Natural Language Processing methods as viable in estimating the behavioral state of the interlocutors. Overview papers [5,6] describe some of those advances. In couples therapy, example work includes [7–16], while similar work exists in other domains such as motivational interview therapy for addiction and autism [17–21]. These methods however suffer from the simplistic assumption of estimating a single behavioral state from the whole interaction: the inherent underlying assumption is that the human is a behavioral state generator and he/she generates from the same state for the entire duration. On the other hand, previous work has also focused on dynamically modeling the behavioral constructs such as emotions [22, 23] and engagement [24]. Whereas these capture the evolution of behavioral states, they fail to provide a link to an overall perception.

This work is the first step towards the creation of a more realistic model of human interaction that will try to address the following: (1) Human annotators employ a range of information to reach an integrated conclusion as to the behavior of interlocutors. While in past work we assumed that all observations (say talk-turns) belonged to the same class as the overall rating, we will expand our models to allow for the human subject to move through a range of behavioral states. We hypothesize that this more realistic model can both (a) provide more accurate overall judgment and (b) higher-resolution information about what happens in the interaction.

(2) Humans do not integrate information in a linear manner. The saying "first impression counts" is an indicator of this non-linear process. This phenomenon is also studied extensively in the psychology literature. For example [25] investigates recency (thing observed last counts more) and primacy (thing observed first counts more). Integrating information requires an in-depth investigation of the algorithmic metrics of behavior (say log likelihood of a negative statement) versus the impact this has on the coding process. In our recent work we briefly touched on this [26] in investigating how human annotators employ isolated saliency or causal integration to make their decisions. Through the higher-resolution information achieved through (1) we can evaluate non-linear techniques in fusing local decisions for reaching a global (session) decision.

This paper provides a proof of concept of the above by employing a 2-state model described below as *Dynamic Behavioral Model* (DBM) and contrasting it with the *Static Behavior Model* (SBM). The SBM works as our baseline assuming a constant behavioral state throughout interaction. A DBM allows for transitions between behavioral states within a session, and learns behavior representations from these transitions. Comparison of these models helps provide evaluation results for the goal (1a) in this work while we are working with our psychology partners to evaluate (1b). In order to evaluate fusion of turn-level decisions as stated in goal (2), we evaluate methods such as model log-likelihood comparison and comparing distribution of behavioral state occupancies.

This paper is organized as follows: Section 2 provides a description of the database. We describe the SBM and DBMs in detail in Section 3, followed by the training methodology in section 4. The evaluation of our behavioral models and their results are detailed in Section 5, after which we discuss the results in Section 6. Finally, we present our conclusions in Section 7.

2. DATABASE

We use the data of 134 couples from the UCLA/UW Couple Therapy Research Project [27]. The corpus consists of audio and video, recorded during sessions of real couples interacting, and the session transcripts. During a session, the husband and wife converse about a pre-decided topic (e.g. "why can't you leave my stuff alone ?") for a certain duration of time. Based on the Couples Interaction Rating System [1] and Social Support Interaction Rating System [2], each participant is later rated by annotators for a set of 33 behavioral codes including "Acceptance", "Blame", "Positivity" and "Negativity". Annotators provide subjective ratings on a Likert scale of 1-9,
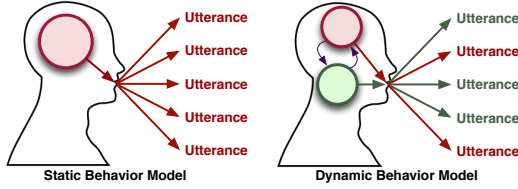
Fig. 1: Conceptualizing the proposed graphical model on the right versus the baseline on the left

where 1 indicates absence of the behavior and 9 implies a strong presence. The sessions are rated by 2-12 annotators with majority of the sessions ($\sim$ 90%) rated by 3-4 annotators.

For our task, we analyze the 'Negativity' behavioral code. In order to simplify the code learning, we only use sessions with mean annotator ratings in the top 20% ('High Negativity') and bottom 20% ('Low Negativity') of the code range and we binarize into C_1 and C_0 respectively. In this manner, we pose the learning problem as a binary classification one, as was also done in our earlier work [16]. Table 1 lists the statistics on the chosen set of data and we describe our modeling schemes in the next section.

For more information, the reader can refer to [1, 2, 10]

	Husband		Wife	
Class Label	C_0	C_1	C_0	C_1
Code Range	1.00-1.75	5.33-9.00	1.00-2.00	6.33-9.00
No. of Files	70	70	70	70

Table 1: Data Demographics on 20% least/most negativity sessions

3. BEHAVIORAL MODELING

Existing work [12] has assumed that an interlocutor is constantly in the same behavioral state. This is equivalent to modeling each inter-locutor in the interaction as a single state generative model as shown in Fig. 1 (left). This is clearly limiting and does not reflect human behavior, that dynamically adapts based on the various stimuli, internal and external. Below we present two models: the *Static Behavior Model* in Sec. 3.1 as our baseline and similar to the one in [12], and the Dynamic Behavior Model in Sec. 3.2 that allows for transition between two behavioral states throughout the interaction and thus makes turn-level decisions instead of session level decisions.

3.1. Static Behavior Models

In the *Static Behavior Modeling* (SBM) framework, a person's behavior is assumed to remain the same throughout the interaction, irrespective of external stimuli such as spouse's utterances, topic being discussed, etc. This is the same model we had proposed in [12] and in effect corresponds to behavioral averaging. Thus, all the utterances observed in that session are generated from the same behavioral state as in Fig. 1(left).

Identifying the behavioral state of the interlocutor in this case is equivalent to identifying the class label $C_i = \{C_0, C_1\}$ from the whole transcript. In this work we employ a *Maximum Likelihood* (ML) formulation for the binary classification. For a set of observed utterances, corresponding to the whole transcript, $\bar{U} = \{U(1), \ldots, U(M)\}$, we want to find:

$$P(\text{Code Low-Negativity or High-Negativity}|\bar{U}) = P(C_0 \text{ or } C_1|\bar{U})$$

For class $C_i = \{C_0, C_1\}$:

$$C_i = \arg\max_{C_j} \frac{P(\bar{U}|C_j)P(C_j)}{P(\bar{U})} \quad (1)$$

$$= \arg\max_{C_j} P(\bar{U}|C_j)P(C_j) \quad (2)$$

Since our dataset is balanced (70 files per class), the class prior probabilities are equal, i.e. $P(C_0)=P(C_1)$. Therefore, we re-write the decision as:

$$C_i = \arg\max_{C_j} P(\bar{U}|C_j) \quad (3)$$

Equation 3 represents the final decision scheme for the SBM. The training and testing methodologies for this model are discussed in Sections 4.1 and 5.1 respectively.

3.2. Dynamic Behavior Models

Dynamic Behavior Models (DBMs) allow for a person's behavior to change over time. This is modeled in the form of transitions between different behavioral states throughout a session, shown in Fig. 1 (right). In our work we will simplify the model to have 2 states and we assume that behavior remains short-term stationary (i.e. behavior does not change within an utterance, but only from one utterance to another).

We will denote utterance states as $S_i = \{S_0, S_1\}$. The behavioral state of the interlocutor, labeled by the human annotators as Low/High Negativity or C_0/C_1, does not provide a one-to-one correspondence any more. A person that is very negative (C_1) can generate from both states (S_0, S_1). The definition of the states S_0, S_1 will be described in detail in Sec. 4.2.

Given only one turn, similar to the formulation of the SBM, a ML model will result in:

$$P(S_i|U(m)) = P(U(m)|S_i)P(S_i) \quad (4)$$

This method estimates turn-level state probabilities. Human coders integrate behavioral information and give a summative opinion on the session level. In order to reach inference from the DBM for session level behavioral descriptors, we need to also integrate such turn-level behavioral information. Below we provide two fusion methods for behavioral integration.

3.3. Fusion Methods for DBMs

There are many perception inspired methods for behavioral integration. For instance in our work [26] we evaluated whether we could judge global behavior based on a locally isolated, yet highly informative event or whether integrating information over time was more effective. The premise of such work is that the human perception process is capable of integrating local events to generate an overall impression at the global level, but this process is not transparent and as such is difficult to replicate. What we can instead do is use local information to derive the same global decisions.

In this work we will employ two fusion methods for the DBM: Activation-based and Likelihood-based.

3.3.1. Activation-based

In the *Activation-based DBM* (ADBM), shown in Fig. 2 (left), local, state-label decisions are made for S_i as shown below, based on (4)

$$S_i = \arg\max_{S_j} P(U(m)|S_j)P(S_j) \quad (5)$$

This fusion method uses a majority vote assumption, where the behavioral state which generates the largest proportion of utterances determines the session label decision. From the training data we can learn the mapping of $S_i \rightarrow C_j, \forall \ i, j \in \{0, 1\}$ as explained in Sec.4.3.1
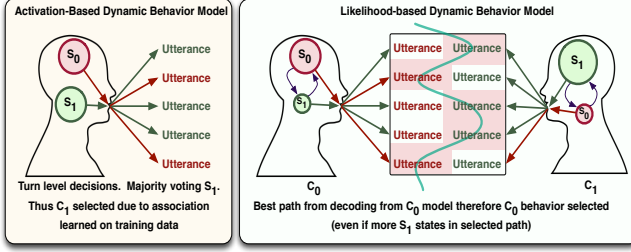
Fig. 2: Activation-based DBM (left) versus the Likelihood-based DBM (right)

3.3.2. Likelihood-based

In the *Likelihood-based DBM* (LDBM) we employ a Hidden Markov Model (HMM) representation of the behavioral classes C_0 and C_1. This model assumes that an interaction is comprised of utterances $\overline{U}=\{U(1),...,U(M)\}$ which are observations emitted by the hidden state sequence $\overline{S}=\{S(1),...,S(M)\}$, where $S(i) \in \{S_0, S_1\}, i \in \{1,...,M\}$. It is also assumed that both classes C_0 and C_1 can produce both states, albeit with different likelihood, as shown in Fig. 2 (right). In human terms, negative persons are likely to remain mostly in negative states and express themselves negatively, although it is sometimes likely that they will express a positive attitude.

In order to identify the most likely behavioral class, we decode $\overline{U}$ based on the HMMs of C_0 and C_1 and chose the model that is more likely to have generated it from the state sequence $\overline{S}$.

$$C_i = \arg\max_{\alpha^j} P(\overline{S}_j|\overline{U}, \alpha^j, \pi) \tag{6}$$

Where
α^j is the HMM state transition matrix of C_j
π is the initial state probability vector
$\overline{S}_j$ is state sequence decoded by HMM of C_j for $\overline{U}$

4. BEHAVIOR MODEL TRAINING

In this section, we describe the training procedure for the SBM and the two DBMs discussed in the previous section. For our implementation, we use language models to represent probabilistic models of lexical content, and build them using the SRILM toolkit [28]. We replace maximum likelihood schemes with minimum perplexity schemes wherever applicable. Perplexity is a measure of how well a probability model predicts an observation; lower the value, better the model. For an utterance $U(m)$, it is calculated as:

$$PP(U(m)) = P(w_1 w_2 ... w_N)^{-1/N} \tag{7}$$

Where
$P(w_1 w_2 ... w_N)$ is joint probability of occurrence of words

---

Algorithm 1 Hard-Assignment EM algorithm for state convergence in activation-based DBM

Initialize utterances in C_0 session $\in \{S_0\}$, $C_1 \in \{S_1\}$
Build language model L_0 from utterances $\in S_0$, L_1 from utterances $\in S_1$
while training perplexity does not converge **do**
    E-step: Classify every utterance $U(m)$ in C_0,C_1 classes
    Get perplexities PP_0, PP_1 of $U(m)$ computed by L_0, L_1
    $$PP_0\{U(m)\} \underset{state=S_0}{\overset{state=S_1}{\gtrless}} PP_1\{U(m)\}; m \in \{0, 1, ..., M\}$$
    M-step: Build L_0 from S_0 utterances,& L_1 from S_1
end while

---

w_i is the ith word of utterance $U(m), i \in \{1, 2, ..., N\}$
From (7) we see that minimizing perplexity is equivalent to maximizing probability

4.1. SBM

From the SBM assumption, we know that the classes and states are equivalent. Therefore, for building a model of a particular behavioral state C_0/C_1, we collect all utterances from the corresponding sessions and train the language models L_0/L_1 respectively. While doing so, we combine our trained LMs with a *Universal Background Model* (UBM), in order to smooth the language models. We use a parameter $\lambda=0.1$ as the interpolation weight for the UBM. Thus, the SBM consists of two LMs for each speaker that model his/her language structure corresponding to the least and the most negative behavior.

4.2. DBM

As seen in Section 3.2, DBMs allow multiple transitions within an interaction and as a result, each session consists of utterances of multiple types of behavior. However, we have only the global session ratings C_0 and C_1 and not the utterance-level state labels. Therefore, we use two iterative semi-supervised methods to label turn-level information into the two clusters corresponding to S_0 and S_1, as described in Section 4.3. Due to the non-availability of utterance-level labels, the convergence of these methods is verified indirectly through the overall training perplexity and the testing accuracy.

4.3. Fusion Methods for DBMs

Fusion methods tie local utterance-level decisions to the global session-level behavior. Therefore, the type of fusion used defines the relation between behavioral states S_0/S_1 and behavioral classes C_0/C_1, as explained below:

4.3.1. ADBM

Just like in the SBM, we initialize the language models L_0/L_1 for states S_0/S_1 from utterances in C_0/C_1 respectively, smoothed with a UBM. We then classify each utterance as S_0 or S_1 and re-train language models until the training perplexity converges. Since perplexities are not direct measures of log-likelihoods, we classify each ut-

---

Algorithm 2 Viterbi-EM algorithm for state and class parameter convergence in likelihood-based DBM

Initialize utterances in C_0 session $\in \{S_0\}$, $C_1 \in \{S_1\}$
Build language model L_0 from utterances $\in S_0$, L_1 from utterances $\in S_1$
Initialize π, α^0, α^1
while training perplexity does not converge **do**
    E-step: Decode C_0 utterances using α^0, L_0, L_1, π
    for every session utterance $U(m)$ **do**
        Get probabilities P_0, P_1 of utterance $U(m)$ from L_0, L_1
        Find probability that $U(m)$ was generated by state S_k
        if m=1 (start of session) **then**
            $\gamma_k(m) = \pi_k * P_k\{U(m)\}; k \in \{0, 1\}$
        else
            $\gamma_k(m) = \arg\max_j [\gamma_j(m-1) * \alpha^0(j,k)] * P_k\{U(m)\}; j,k \in \{0, 1\}$
        end if
        $$\gamma_0(m) \underset{state=S_1}{\overset{state=S_0}{\gtrless}} \gamma_1(m)$$
    end for
    Repeat E-step for C_1 utterances; replace $\boldsymbol{\alpha^0}$ with $\boldsymbol{\alpha^1}$
    M-step: Re-estimate states, class parameters
    Build L_k from all $U(m)$ whose $state = S_k; k \in \{0, 1\}$
    Update $\alpha^n(i,j) = \frac{count(<i,j> \text{ state pairs in class } C_n)}{\sum_k count(<i,k> \text{ state pairs in class } C_n)}$; i,j,k $\in \{0, 1\}, n \in \{0, 1\}$
end while

---

terance independently of the rest, for which a Hard-Assignment Expectation Maximization (EM) scheme is better suited, as described in Algorithm 1. At the completion of this re-assignment, each class now contains both states. We, therefore, compute the proportions of state occupancies for C_0 and C_1 sessions as decoded using the models L_0 and L_1.

4.3.2. LDBM

We initialize our model parameters as in Section 4.3.1 and obtain converged estimates of behavioral states and the class parameters, using the Viterbi-EM algorithm, as described in Algorithm 2. In this model, the Hidden Markov Model (HMM) parameters need to be re-estimated over iterations. Finally, each behavioral state is described by the initial vector π, which contains the probability of the spouse starting in each state, and its language model. Each class is now associated with a matrix that governs state transitions in that class; α^0 for the C_0 sessions, and α^1 for C_1. $\alpha^0(i, j)$ represents the probability of a C_0-rated spouse transitioning to state j, given that he/she was previously in state i.

5. BEHAVIORAL MODEL EVALUATION

In this section, we explain our evaluation procedure for the two different models and the two different fusion techniques used in DBMs.

5.1. SBM

For a given test session, we assign C_1/C_0 label to a speaker after comparing the perplexities from his/her LMs. Given a set of M utterances $\overline{U}=\{U(1), ..., U(M)\}$ from a speaker during the test session, we compute the LM perplexities based on L_0 and L_1, and the class label assignment is shown in equation 8.

$$C_i = \arg\min_j \sum_m PP_j\{U(m)\} \qquad (8)$$

Where
$PP_i\{\}$ represents perplexity score of utterance computed by LM L_i
The results for SBM, obtained using a 1-gram LM, are shown in Table 2.

5.2. DBM

Although both ADBM and LDBM allow state transitions, the scoring mechanism is quite different in each case, as explained below.

5.2.1. ADBM

For a given test session, we perform hard-assignment of state labels to utterances using L_0 and L_1 and compute the state distribution. For a set of utterances $\overline{U}=\{U(1), ..., U(M)\}$ where the most commonly occurring state is found to be S_k, the ADBM chooses the behavioral class that maximizes

$$C_i = \arg\max_{C_j} P(S_k|C_j) \qquad (9)$$

The performance of ADBM improves, on average, by around 1% relative to the SBM, and the results are shown in Table 2.

5.2.2. LDBM

For a set of test utterances $\overline{U}$, we decode based on HMMs of both C_0 and C_1. The label of the class whose prediction is associated with the highest likelihood is assigned to the test file. This decision scheme is shown in equation (10)

$$C_i = \arg\max_{\lambda_j} P(\overline{S}_j|\overline{U}, \alpha^j, \pi) \qquad (10)$$

Where
$\overline{U}$ is set of test utterances, $\{U(1), ..., U(M)\}$
$\overline{S}_n$ is the test state sequence predicted by class C_n, $\{S(1), ..., S(M)\}$
λ_n is set of HMM parameters of class C_n, $\{\alpha^n, L_0, L_1, \pi\}$
The performance of LDBM improves, on average, by around 5.7% relative to the SBM, and the results are shown in Table 2.

Model	Husband	Wife	Average
SBM	79.29%	83.57%	81.43%
Activation-DBM	85.00%	79.29%	82.15%
Likelihood-DBM	83.57%	88.57%	86.07%

Table 2: Classification Accuracy of Behavioral Models using 1-grams

6. DISCUSSION

The likelihood-DBM has the best average performance across both spouses, at 86.07%, while the SBM has the lowest, with 81.43%. This matches our expectations since the likelihood-DBM is less rigid in its assumptions about behavioral changes. Thus, it is better equipped to capture information about dynamic behavior, as compared to the other two models.

While the SBM helps in identifying which state a person's behavior more likely corresponded to, it is limited in its ability to model changes in behavior within a session. The Activation-DBM can model behavioral changes within the same session, and performs well, but it allows rapid changes in behavioral states across successive utterances. We feel that such fast changes in behavior do not realistically mirror the usual changes in human behavioral expression. In addition, there is no dependence on previous utterances, meaning that context is ignored. The Likelihood-DBM avoids the pitfalls associated with the first two models, and while more complex, it is more accurate in predicting behavior from language. Note that both dynamic implementations can likely be improved through supervised learning and it is notable that the achieved performance is through only loose semi-supervised clustering.

7. CONCLUSION

Automating the psychological evaluations can have a significant impact by providing less expensive and more accurate methods of behavioral coding. In this work, we proposed a Dynamic Behavior Model based scheme which models a spouse in couples' therapy as transitioning through multiple behavioral states to obtain an overall perception of negativity. We tested two models for dynamically modeling the behavior: Activation-based DBM and Likelihood-based DBM which outperformed the Static Behavioral Model. Furthermore, the LDBM performed slightly better than ADBM as it provided more freedom to the way behavior could be expressed.

In this work, we addressed only two models of local information integration towards deriving global behavioral descriptions. We plan to extend this model to further incorporate dependencies amongst spouses to model their behavioral influences as well as dependencies among other codes apart from negativity. So far, we have only considered two state models to predict binary behavior levels. We will develop models with a finer behavior stratification and observe its relation with the number of states. Finally, this model can be employed in a range of application domains involving behavioral evaluations in multi-party interactions.

8. REFERENCES

[1] C. Heavey, D. Gill, and A. Christensen, *Couples interaction rating system 2 (CIRS2)*, University of California, Los Angeles, 2002.

[2] J. Jones and A. Christensen, "Couples interaction study: Social support interaction rating system," Technical manual, University of California, Los Angeles, 1998.

[3] RE Heyman and D. Vivian, "Rmics: Rapid marital interaction coding system: Training manual for coders, state university of new york, stony brook, ny (1993)," *Unpublished manuscript. Available at*, 1993.

[4] G. Margolin, P.H. Oliver, E.B. Gordis, H.G. O'Hearn, A.M. Medina, C.M. Ghosh, and L. Morland, "The nuts and bolts of behavioral observation of marital and family interaction," *Clinical Child and Family Psychology Review*, vol. 1, no. 4, pp. 195–213, 1998.

[5] Shrikanth S. Narayanan and Panayiotis G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceeding of the IEEE*, 2014.

[6] Panayiotis G. Georgiou, Matthew P. Black, and Shrikanth S. Narayanan, "Behavioral signal processing for understanding (distressed) dyadic interactions: some recent developments," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, New York, NY, 2011, J-HGBU '11, pp. 7–12, ACM.

[7] Bo Xiao, Panayiotis Georgiou, and S. Narayanan, "Data driven modeling of head motion towards analysis of behaviors in couple interactions," in *International Conference on Audio, Speech and Signal Processing*, 2013.

[8] James Gibson, Bo Xiao, Panayiotis Georgiou, and S. Narayanan, "An audio-visual approach to learning salient behaviors in couples' problem solving discussions," in *International Conference on Audio, Speech and Signal Processing*, 2013.

[9] Chi-Chun Lee, Athanasios Katsamanis, Matthew P. Black, Brian Baucom, Andrew Christensen, Panayiotis G. Georgiou, and Shrikanth S. Narayanan, "Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions," *Computer, Speech, and Language*, 2012.

[10] M.P. Black, A. Katsamanis, B.R. Baucom, C.C. Lee, A.C. Lammert, A. Christensen, P.G. Georgiou, and S.S. Narayanan, "Toward automating a human behavioral coding system for married couples interactions using speech acoustic features," *Speech Communication*, 2012.

[11] Matthew Black, Panayiotis Georgiou, Athanasios Katsamanis, Brian Baucom, and Shrikanth Narayanan, ""You made me do it": Classification of blame in married couples' interaction by fusing automatically derived speech and language information," in *Proceedings of InterSpeech*, Florence, Italy, August 2011.

[12] P. G. Georgiou, M. P. Black, A. Lammert, B. Baucom, and S. S. Narayanan, ""That's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?," in *Proceedings of Affective Computing and Intelligent Interaction*, Memphis, TN, USA, 2011.

[13] C.C. Lee, A. Katsamanis, M. Black, B. Baucom, P. Georgiou, and S. Narayanan, "Affective state recognition in married couples interactions using PCA-based vocal entrainment measures with multiple instance learning," *Affective Computing and Intelligent Interaction*, pp. 31–41, 2011.

[14] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, P. G. Georgiou, and S. S. Narayanan, "An analysis of PCA-based vocal entrainment measures in married couples' affective spoken interactions," in *Proceedings of InterSpeech*, Florence, Italy, 2011.

[15] C.-C. Lee, M. P. Black, A. Katsamanis, A. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. Narayanan, "Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples," in *Proceedings of InterSpeech*, 2010.

[16] M. P. Black, A. Katsamanis, C.-C. Lee, A. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Automatic classification of married couples' behavior using audio features," in *Proceedings of InterSpeech*, 2010.

[17] Bo Xiao, Daniel Bone, Maarten Van Segbroeck, Zac E. Imel, David Atkins, Panayiotis Georgiou, and Shrikanth Narayanan, "Modeling therapist empathy through prosody in drug addiction counseling," in *In Proceedings of Interspeech*, 2014.

[18] Daniel Bone, Chi-Chun Lee, Theodora Chaspari, Matthew P. Black, Marian Williams, Sungbok Lee, Pat Levitt, and Shrikanth S. Narayanan, "Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand," in *Proceedings of InterSpeech*, Aug. 2013.

[19] Theodora Chaspari, Chi-Chun Lee, and Shrikanth Narayanan, "Interplay between verbal response latency and physiology of children with autism during ECA interactions," in *Proceedings of InterSpeech*, 2012.

[20] Theodora Chaspari, Emily Mower Provost, Athanasios Katsamanis, and Shrikanth Narayanan, "An acoustic analysis of shared enjoyment in ECA interactions of children with autism," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, 2012.

[21] E. Mower, C.C. Lee, J. Gibson, T. Chaspari, M.E. Williams, and S. Narayanan, "Analyzing the nature of ECA interactions in children with autism," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[22] Rahul Gupta, Nikolaos Malandrakis, Bo Xiao, Tanaya Guha, Maarten Van Segbroeck, Matthew P Black, Alexandros Potamianos, and Shrikanth S Narayanan, "Multimodal prediction of affective dimensions and depression in human-computer interactions," .

[23] Björn Schuller, Gerhard Rigoll, and Manfred Lang, "Hidden markov model-based speech emotion recognition," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. IEEE, 2003, vol. 2, pp. II–1.

[24] Gian-Marco Baschera, Alberto Giovanni Busetto, Severin Klingler, Joachim M Buhmann, and Markus Gross, "Modeling engagement dynamics in spelling learning," in *Artificial Intelligence in Education*. Springer, 2011, pp. 31–38.

[25] Dirk D Steiner and Jeffrey S Rain, "Immediate and delayed primacy and recency effects in performance evaluation.," *Journal of Applied Psychology*, vol. 74, no. 1, pp. 136, 1989.

[26] Chi-Chun Lee, Athanasios Katsamanis, Panayiotis G. Georgiou, and Shrikanth S. Narayanan, "Based on isolated saliency or causal integration? toward a better understanding of human annotation process using multiple instance learning and sequential probability ratio test," in *Proceedings of InterSpeech*, Sept. 2012.

[27] A. Christensen, D.C. Atkins, S. Berns, J. Wheeler, D.H. Baucom, and L.E. Simpson, "Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples," *Journal of Consulting and Clinical Psychology*, vol. 72, no. 2, pp. 176–191, 2004.

[28] Andreas Stolcke et al., "Srilm-an extensible language modeling toolkit.," in *INTERSPEECH*, 2002.