UNSUPERVISED LATENT BEHAVIOR MANIFOLD LEARNING FROM ACOUSTIC FEATURES: AUDIO2BEHAVIOR

Haoqi Li[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

Behavioral annotation using signal processing and machine learning is highly dependent on training data and manual annotations of behavioral labels. Previous studies have shown that speech information encodes significant behavioral information and be used in a variety of automated behavior recognition tasks. However, extracting behavior information from speech is still a difficult task due to the sparseness of training data coupled with the complex, high-dimensionality of speech, and the complex and multiple information streams it encodes. In this work we exploit the slow varying properties of human behavior. We hypothesize that nearby segments of speech share the same behavioral context and hence share a similar underlying representation in a latent space. Specifically, we propose a Deep Neural Network (DNN) model to connect behavioral context and derive the behavioral manifold in an unsupervised manner. We evaluate the proposed manifold in the couples therapy domain and also provide examples from publicly available data (e.g. standup comedy). We further investigate training within the couples' therapy domain and from movie data. The results are extremely encouraging and promise improved behavioral quantification in an unsupervised manner and warrants further investigation in a range of applications.

Index Terms— Behavior Signal Processing, manifold learning, unsupervised learning, behavior representation

1. INTRODUCTION

Analysis and classification of human behaviors is one of the core tasks of observational study. For example, in couples therapy, psychologists observe and identify domain-specific behaviors(*e.g.,* blame and acceptance) during couple interactions, and provide specific treatments based on their analysis.

Behavior estimation process is a complicated task. Different from emotions, human behaviors such as acceptance, are often manifested over long time scales. Longer context needs to be considered when human annotators attempt to quantify behavior. Because of that, human raters need to combine information at different timescales to estimate behaviors correctly. It is difficult to simulate the complex non-linear nature of the annotation process using one specific algorithm. Moreover, data with rich behavioral information from psychotherapy domains are often severely limited in quantity due to privacy constraints and cost of annotation.

Integrating machine learning and signal processing methods, Behavior Signal Processing (BSP)[1, 2] employs acoustic[3, 4], lexical[5, 6], and visual[7, 8] information to model and analyze multi-modal human behaviors. For example, in couples therapy domain, using acoustic features, Black *et al.*[3] built an automatic human behavioral coding system for couples interaction. To deal with data sparsity, a sparsely connected and disjointly trained deep neural networks (SD-DNN) framework was introduced in [9], that limits the number of trained parameters at any time.

Despite these efforts in the BSP domain, it is still challenging to extract effective behavior representations from high-dimensionality acoustic features. Over the last few years, Deep Neural Networks have demonstrated promise in their capability to learn high level representation from raw data. For instance, by training DNN with audio features input, and corresponding labels(*e.g.,* emotion recognition in [10, 11], keyword spotting in [12]) as target, the output of DNN can be regarded as representation of raw input data. However, this supervised framework fails in our specific domain, since a huge amount of training data with annotated labels is essential. Data sparseness limits the use of AI methods for emotions, stress, and behavior estimation. Thus, in this work we propose an unsupervised way of exploiting data for the BSP domain. We further investigate whether out of domain data can be employed for in-domain behavioral quantification.

Recently, context information has been used for a range of applications. For instance in developing the word2vec model Milikov *et al.*[13, 14] have proposed an embedding that ties 1-hot word representations of nearby words via an intermediate, hidden, vector representation. Similar to auto-encoders or bottleneck representations[15, 16], the hidden layer attempts to connect the information at the input and output layers, but in this case the information resides at a longer scale than either of the two representations – namely context.

Our proposed framework employs a similar idea to the word2vec. Since humans employ a large temporal window to observe the context and evaluate behaviors, we can hence assume that behavior remains relatively constant within a sufficiently long window. This matches also annotation guidelines in the field of psychology where the minimum observation windows are usually set at 30 seconds. It also matches empirical understanding of behavior. For example, one person (often the case in couples therapy interactions as well as everyday life) can be sad during a conversion for a long window of time despite different intonations and speech patterns throughout that temporal window.

In our paper, we propose an unsupervised behavior manifold learning using Deep Neural Network via unlabeled acoustic features. We learn the manifold with unlabeled within-domain data and from Out-Of-Domain (OOD) data. We evaluate if the knowledge gained includes behaviorally meaningful information within and OOD training and within and OOD testing.

The rest of paper is organized as follows: Section 2 describes in detail our proposed manifold learning to obtain behavior representation in an unsupervised manner. Section 3 provides a brief description of the database used in our paper, after which we describe

Work supported by NSF and DoD.

audio processing, feature extraction steps and experiment settings in section 4. After that, we discuss our results in section 5. Finally, we give our conclusion and future work in section 6.

2. METHODOLOGY

The success of machine learning algorithms can be attributed to two main properties: first the DNN can represent any function, and second it can learn that function based on large amounts of data. The underlying representations that the DNN identifies are critical to its success[17, 18]. In the BSP domain, we often suffer from lack of data while the complexities of the signal require the use of high-dimensionality acoustic features. The goal of this paper is to identify, in an unsupervised manner, a latent manifold where the signal retains its behavioral characteristics. In this behavioral manifold we expect similar behaviors to appear closer together than they do in the original signal space or in the feature space. Based on the geometric notion of manifolds, the learned representation can be associated with an intrinsic coordinate system on the embedded manifold[17]. In our case, an effective behavior manifold should preserve information residing on a "behavioral axis", while removing other acoustically encoded information.

One reasonable assumption is that the behavioral state of a person is slow varying (note that behavior changes much slower than emotional expression despite the close relations between the two). This means that by looking at a very short interval of behavior (say 5s) and a following interval (say next 5s), we will most likely observe the same or a very similar behavioral state. Based on this assumption we will create a model that exploits context and ties the two intervals via the proposed reduced dimensionality embedding vector space.

We acknowledge and expect the following complication with the above assumption: the nearby information frames also encode speaker characteristics as well as acoustic conditions such as environment and channel. We will discuss this further in Section 5.

2.1. Training framework

Our proposed training framework is similar to an autoencoder, but rather than just training to reconstruct the input our system trains to reconstruct neigboring frames. As shown in Fig. 1, for the kth frame of acoustic features, the outputs are frames from $k-w$, $k+w$ excluding the kth frame, where w is the size of the window in which we consider behavioral context to remain relatively constant. By creating such an unsupervised corpus we can train similarly to standard DNN tasks with back propagation, thus learning the underlying behavioral manifold representation.

2.2. Behavior manifold representation

After the training, we use the output of the bottleneck layer as the behavior representation. In general, the dimension of the hidden layer is smaller than dimension of the original feature space, so this process can be also regarded as a feature dimensionality reduction or compression process.

2.3. Evaluation

Since we employ an unsupervised method in training our model, we need to demonstrate that representations indeed include behavior information. We intend to do this on different evaluation data: (i) From the field of psychology we will employ as a case study Couples Therapy interactions and we will compare underlying representations of similarly rated sessions. For example, after learning manifold on unsupervised data, a test session (with known rating) can be compared in the manifold space with all known samples of negative/positive behaviors, and closest match can be selected. (ii) We also collect a range of samples from political speeches, standup comedy *etc.*and compare their pairwise similarity. Details of the datasets are provided below.

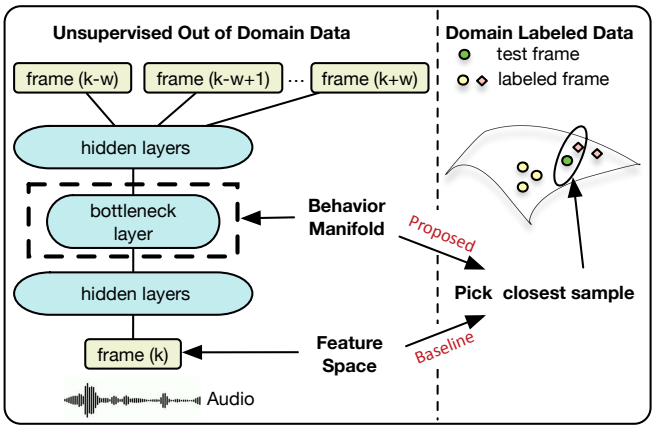

Fig. 1. Behavior representation training framework

3. CORPUS

For the unsupervised training process we utilize two corpora in our paper:

- *Ti:* For in-domain BSP data (Train-in-domain: Ti) we employ the couples therapy database by UCLA/UW Couple Therapy Research Project[19], in which 134 couples were involved in video-taped marital issue interactions. In each session, one relationship-related topic (*e.g.,* "Why can't you leave my stuff alone?") was initiated during the speech session. Although not used for the training, behavioral labels exist for this corpus.
- *To:* For out of BSP domain training dataset, we collected around 400 hours of audio from a range of movies. Many of the selected movies include large parts of emotional conversions reflecting a range of behaviors.

For testing we also employ two datasets:

- *Eo:* For out of domain evaluation (Eo) data, we collected audio from two different speakers for each of the following scenarios: stand-up comedy of commedians who employ anger as an elicitation mechanism (see Table 1), comedian without angry behavior, political debate, Ted talk, eulogy. Each audio's length is around 10 minutes.
- *Ei:* Within the BSP domain we employ the labels of our couples therapy data. Each participant's behavior was evaluated by trained human annotators for a set of 33 behaviors(*e.g.,* "Acceptance", "Blame" *etc.*) based on standard Couples Interaction[20] and Social Support Rating Systems[21]. Each annotator rated 1-9 for each behavior at session level in terms of the presence of this behavior. In this work we show relationships with 4 of the behaviors by binarizing the top and bottom 20% of the original ratings.

Table 1. Out of Domain test data

1. George Carlin; 2. Richard Pryor; 3. Jim Gaffigan; 4. Steve Hofstetter 5. Final Republican Presidential Debate, 2015 6. Vice Presidential Debate 2012 7. TEDtalk: Kevin Slavin; 8. Christopher Steiner 9. Eulogy for a Son (youtube) 10. Mr. Li Hongyi's Eulogy for the late Mr. Lee Kuan Yew

4. EXPERIMENTS

4.1. Audio processing

For couples therapy data, due to the limitations of the available recordings, some pre-processing is needed to remove sessions with low SNR. Further, since these are dyadic interaction data, we wanted to diarize the interactions. In this work, we employed the same pre-processing as in[3]. In short, we utilize all interactions with an SNR above 5dB, perform Voice Activity Detection (VAD) to identify spoken regions, and Speaker Diarization to identify same-speaker regions.

For the movie dataset we did not perform any pre-processing procedure, thus treating all frames the same, including silence, music, and changing speaker regions.

4.2. Acoustic feature extraction

We extract acoustic features characterizing speech prosody (pitch, intensity and their derivatives), spectral envelope characteristics (MFCCs, MFBs, LPCs and their dirivatives), voice quality (jitter, shimmer and their dirivatives). All of these Low-Level Descriptors (LLD) are extracted every 10ms with a 25ms Hamming window using openSMILE[22]. Within each frame, we compute functionals of all these acoustic features including Min (1st percentile), Max (99th percentile), Range (99th percentile – 1st percentile), Mean, Median, and Standard Deviation.

Temporal variation of behavior is much slower than basic emotion, thus a longer size of frame window is necessary for its analysis. In our paper, in order to estimate meaningful behavioral metrics while maintaining high resolution, we use a 20s and 5s windows with 1s shift.

4.3. Experimental Setup

We conducted three different types of experiments, using different training or evaluation datasets. In our experiments, the input dimension of our feature is 420 as discussion in section 4.2. The dimension of the bottleneck layer is set to be 64. To generate the training data, for each frame, the window range w is set to be 6, we randomly pick up 5 context frames within context window as reconstruction labels for this frame. For instance for an input of audio from 100-120s the context could be any five frames from $(100 + i) - (120 + i)$ where $i = [-6, -5, \ldots, 6]$.

Three experiments are described as follows:

- **Experiment (1):** Unsupervised training on couples therapy corpus (Ti), and evaluate on Couples therapy corpus (Ei).
- **Experiment (2):** Unsupervised training on movie corpus (To), and test on Couples Therapy corpus (Ei).
- **Experiment (3):** Unsupervised training on movie corpus (To), and test on audio sessions listed in table 1 representing different behavior styles (To, Eo).

4.4. Evaluation Method for In-Domain Couples Therapy

As mentioned before, we have only session-level ratings for the couples therapy corpus. For each behavior code and each gender, we selected 70 sessions on one extreme case of this code (*e.g.,* high acceptance) and another 70 sessions at the other extreme (*e.g.,* low acceptance). We binarize the behavior to provide evaluation class labels and achieve higher inter-annotator agreement. For the couples therapy dataset we will use a supervised evaluation procedure, even though the behavioral manifold has been trained in an unsupervised manner.

After we obtain the latent manifold representation for each frame, we use the Euclidean distance to find the closest "reference frame", which is the nearest frame among all the labeled frames from different couples. The leave-one-couple-out test procedure can ensure a fair evaluation where the speaker characteristics will not have any impact during testing. Further, we use the corresponding session behavior label as the reference frame's label. Then, we employ majority voting to generate session level labels from multiple frame level labels.

4.5. Evaluation Method for OOD data

Unlike the couples therapy data, our OOD do not have any labels. The evaluation data was selected however to reflect different behavioral styles. For instance as seen from Table 1, a politician speaking during a debate is expected to be very different in behavioral style from a stand-up comedian, but similar to another politician. In this case we present the results of the clustering: what frame was close to what as a percentage. This percentage score implies the similarity between two audio frames.

5. RESULTS AND DISCUSSION

5.1. Testing on within domain corpus

Baseline of couples' behavior classification As a baseline system we use a nearest neigbor behavior classification in the acoustic feature space at the frame level, and similarly to 4.4 use majority voting to generate session level labels. The results of this baseline classification method are shown in both Table 2 and Table 3, which are only slightly better than random guess. This result suggests that original acoustic features are not an effective candidate for behavior representation. Further training is needed in order to extract behavior information from high dimensional acoustic features.

Comparison of within and OOD training In order to compare within and OOD training, we conduct experiment (1) and (2) on behavior code *Acceptance*. To be consistent with our precious work[4, 9], a 20s frame size is chosen. Because of limited in-domain dataset (Ti) size, we build a neural network with only 2 hidden layers in that case. When training with out-of-domain data (To), since more training data is available, we employ a neural network with 5 hidden layers of 300, 200, 64, 200, 300 nodes respectively.

From the results in Table 2, both *Ti* and *To* training methods beat the performance of baseline, which shows that our audio2behavior framework is an effective way to project the signal on a more meaningful behavioral manifold in an unsupervised manner and reduce the feature dimensionality. As expected, in-domain training performs better than that from OOD one. This is reasonable, since in terms of speech patterns and acoustic characteristics, there is a big gap between the movie and Couples therapy corpus, and importantly the couples data are far-field, low quality recordings while the movie

Table 2. Classification accuracy (%) of behavior acceptance

Baseline	Train on Couples' 20 s window size	Train on Movies 20 s window size
57.5	69.29	66.43

Table 3. Classification accuracy (%) for behaviors with different frame window size

Behavior	Baseline	Train on Movies **20** s window size	Train on Movies **5** s window size
Acceptance	57.50	66.43	68.57
Blame	55.00	61.07	71.78
Negtivity	63.93	63.93	69.64
Positivity	51.07	65.00	66.43
Average	56.88	64.11	**69.11**

Selected		Comedy		Comedy		Debate		Ted talk		Eulogy	
Input		1	2	3	4	5	6	7	8	9	10
Comedy	1	0.00	**0.44**	0.13	0.19	0.04	0.05	0.05	0.03	0.05	0.02
	2	**0.38**	0.00	0.19	0.13	0.03	0.03	0.04	0.06	0.10	0.05
Comedy	3	0.17	0.24	0.00	**0.24**	0.06	0.04	0.06	0.05	0.09	0.04
	4	**0.31**	0.18	0.18	0.00	0.06	0.02	0.07	0.05	0.07	0.05
Debate	5	0.12	0.08	0.19	0.12	0.00	**0.21**	0.13	0.09	0.05	0.02
	6	0.14	0.08	0.14	0.08	**0.16**	0.00	0.09	0.12	0.10	0.09
Ted talk	7	0.07	0.08	0.13	0.11	0.05	0.06	0.00	**0.31**	0.11	0.08
	8	0.07	0.09	0.08	0.08	0.05	0.07	**0.24**	0.00	0.19	0.12
Eulogy	9	0.08	0.13	0.10	0.09	0.03	0.03	0.10	0.16	0.00	**0.29**
	10	0.08	0.05	0.08	0.08	0.01	0.03	0.05	0.16	**0.47**	0.00

Fig. 2. Behavior scenario similarity evaluation results

data are usually higher quality signals. The mismatch of the training and test sets is minimal when both are from the same domain.

Comparison of frame length The OOD training result in Table 2 is promising, especially for out-of-domain dataset, since we do not perform any pre-processing procedures, such as VAD or diarization. Because of that, as we mentioned in 4.1, there should be some non-speech parts, such as background music, silence, as well as multiple sources of noise besides human speech. In addition, within a larger frame window size, it is also highly probable that speech regions within each window come from multiple speakers. Different speaker's characteristics in one frame window may contaminate behavior related acoustic representation. We try to find a proper way to improve the performance by reducing speaker characteristics. One approach is to reduce the length of frame window. We thus hypothesize that using a smaller window size the chance of single-speaker regions in each frame becomes higher and thus it should improve the audio2behavior model performance by lowering acoustic complexity. This clearly assumes that the window, while smaller, is still long enough to capture the behavioral characteristics.

We employ experiment (2) on multiple behaviors with different frame lengths to verify this hypotheses. From the results in table 3, we can see there is significant improvement, a 5% absolute increase from 64.11% to 69.11% in terms of classification accuracy. Moreover, for all four behaviors, a consistent improvement is noted on 5s frame length acoustic feature. This shows that consistency within each acoustic speech frame region might be one critical issue in audio2behavior system, and encourages diarization as a front end pre-processing step. We should note here that for complex human behavior annotation process, even for human annotators, the inter-annotator agreement can only reach about Krippendorffs $\alpha = 0.8$ [6], and so the 69.11% for a completely unsupervised method with just majority vote at the output is very encouraging.

In general, these results are promising for communicative behavior quantification since we only utilize unlabeled, any-domain data and train in an unsupervised manner.

5.2. Testing on OOD corpus

As mentioned in section 3, we collect OOD test dataset from different scenarios listed in Table 1. In each scenario, two audio files are collected from different speakers. We use normalized percentage score to evaluate behavior similarity. The score is calculated by dividing number of nearest frames in each selected scenario by the number of total frames of input audio. Results are shown in Figure 2.

Ideally, audio from similar scenarios should exhibit high similarity with each other, and a lower score should be assigned between less related scenarios. We find that 9 out of 10 audio samples are classified as we hoped based on majority vote on frame level clustering. Moreover, besides classification, results show behavior similarity with details: audio2behavior can show behavior similarity under different degrees. For example, we can see that 44% of data from George Carlin are identified as similar to Richard Pryor, where both comedians employ an angry tone in their standup comedy and 19% and 13% come from Steve Hofstetter and Jim Gaffigan, also comedians that employ a milder tone in their routines. Less than 5% of the data are associated with any of the other conditions. All these results show promising behavioral quantification of our audio2behavior model.

6. CONCLUSION AND FUTURE WORK

Data sparsity is always a critical issue in behavior related studies. Behavior recognition research suffers from expensive data annotation process and low inter-annotator agreement, which also limits the performance of automated behavior recognition system. Compared with previous existing supervised behavioral recognition in BSP domain, our audio2behavior provides another possible solution candidate: transfer out of domain knowledge into training, then adapt the model into domain applications. This unsupervised training approach of vectorizing abstract behavior from audio and then obtaining better behavioral quantification in manifold shows auspicious results and applications in behavioral signal processing domain.

In the future, inspired by results of this paper, we plan to employ VAD and diarization into the front end to better improve the training of the audio2behavior model. This will reduce speaker characteristics and acoustic complexity in behavior representation by allowing us to do speaker-specific normalizations. Alternatively we can employ the speaker-distinct regions but in a joint and unsupervised manner learn both a speaker and behavioral manifold.

Moreover, unsupervised behavior representation models can be also employed into a range of applications for which training data are unavailable, by quickly allowing out-of-domain bootstrapping.

7. REFERENCES

[1] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, May 2013.

[2] 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, USA, 2011, J-HGBU '11, pp. 7–12, ACM.

[3] Matthew P. Black, Athanasios Katsamanis, Brian R. Baucom, Chi-Chun Lee, Adam C. Lammert, Andrew Christensen, Panayiotis G. Georgiou, and Shrikanth S. Narayanan, "Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features," *Speech Communication*, vol. 55, no. 1, pp. 1 – 21, 2013.

[4] Wei Xia, James Gibson, Bo Xiao, Brian Baucom, and Panayiotis G Georgiou, "A dynamic model for behavioral analysis of couple interactions using acoustic features," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[5] Panayiotis G. Georgiou, Matthew P. Black, Adam Lammert, Brian Baucom, and Shrikanth 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 (ACII), Lecture Notes in Computer Science*, Oct. 2011.

[6] Shao-Yen Tseng, Sandeep Nallan Chakravarthula, Brian Baucom, and Panayiotis Georgiou, "Couples behavior modeling and annotation using low-resource LSTM language models," in *Proceedings of Interspeech*, San Francisco, CA, September 2016.

[7] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 1107–1119, July 2015.

[8] A. Metallinou, R. B. Grossman, and S. Narayanan, "Quantifying atypicality in affective facial expressions of children with autism spectrum disorders," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, July 2013, pp. 1–6.

[9] Haoqi Li, Brian Baucom, and Panayiotis Georgiou, "Sparsely connected and disjointly trained deep neural networks for low resource behavioral annotation: Acoustic classification in couples' therapy," in *Proceedings of Interspeech*, San Francisco, CA, September 2016.

[10] Kun Han, Dong Yu, and Ivan Tashev, "Speech emotion recognition using deep neural network and extreme learning machine.," in *Interspeech*, 2014, pp. 223–227.

[11] D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using hidden markov models with deep belief networks," in *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, Dec 2013, pp. 216–221.

[12] Guoguo Chen, Carolina Parada, and Tara N Sainath, "Query-by-example keyword spotting using long short-term memory networks," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5236–5240.

[13] T Mikolov and J Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, 2013.

[14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[15] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[16] Pierre Baldi, "Autoencoders, unsupervised learning, and deep architectures.," *ICML unsupervised and transfer learning*, vol. 27, no. 37-50, pp. 1, 2012.

[17] Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation Learning: A Review and New Perspectives," *arXiv.org*, June 2012.

[18] Yoshua Bengio et al., "Deep learning of representations for unsupervised and transfer learning.," *ICML Unsupervised and Transfer Learning*, vol. 27, pp. 17–36, 2012.

[19] Andrew Christensen, David C Atkins, Sara Berns, Jennifer Wheeler, Donald H Baucom, and Lorelei 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, 2004.

[20] C Heavey, D Gill, and A Christensen, "Couples interaction rating system 2 (CIRS2)," *University of California, Los Angeles*, vol. 7, 2002.

[21] J Jones and A Christensen, "Couples interaction study: Social support interaction rating system," *University of California, Los Angeles*, vol. 7, 1998.

[22] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM International Conference on Multimedia*, New York, NY, USA, 2013, MM '13, pp. 835–838, ACM.