

Complexity in speech and its relation to emotional bond in therapist-patient interactions during suicide risk assessment interviews

Md Nasir¹, Brian Baucom², Craig J. Bryan², Shrikanth Narayanan¹, Panayiotis Georgiou¹

¹University of Southern California, Los Angeles, CA, USA

²University of Utah, Salt Lake City, UT, USA

mdnasir@usc.edu, {brian.baucom, craig.bryan}@utah.edu, {shri, georgiou}@sipi.usc.edu

Abstract

In this paper, we analyze a 53-hour speech corpus of interactions of Soldiers, who had recently attempted suicide or had strong suicidal ideation, conversing with their therapists. In particular, we study the complexity in therapist-patient speech as a marker of their emotional bond. Emotional bond is the extent to which the patient feels understood by and connected to the therapist. First, we extract speech features from audio recordings of their interactions. Then, we consider the nonlinear time series representation of those features and compute complexity measures based on the Lyapunov coefficient and correlation dimension. For the majority of the subjects, we observe that speech complexity in therapist-patient pairs is higher for the interview sessions, when compared to that of the rest of their interactions (intervention and post-interview follow-up). This indicates that entrainment (adapting to each other's speech) between the patient and the therapist is lower during the interview than regular interactions. This observation is consistent with prior studies in clinical psychology, considering that assessment interviews typically involve the therapist asking routine questions to enquire about the patient's suicidal thoughts and feelings. In addition, we find that complexity is negatively correlated with the patient's perceived emotional bond with the therapist.

Index Terms: nonlinear dynamical systems, complexity, suicide, prosody, behavioral analysis, dyadic conversations

1. Introduction

Assessment of suicide risk is crucial in the mental health domain and has been an important subject of research [1]. In a typical risk assessment setting, the therapist interviews the patient about his or her suicidal thoughts, feelings and past history of relevance in order to understand the patient's intent to commit suicide. From the therapist's standpoint, maintaining a good therapeutic relationship toward a common goal is of utmost importance during the interactions. If the patient feels less connected or comfortable with the therapist, that can potentially lead to failure in accurate risk assessment. Emotional bond is conceptually related to the notion of empathy, the therapist's ability to feel for the patient's sufferings. Empathy is deemed an essential quality for the efficiency of a psychotherapist [2]. In addition to empathy, emotional bond also entails the patient's feeling of trust and acceptance towards the therapist. This construct's bidirectional and interpersonal nature makes it useful yet challenging to analyze from spoken interactions.

Analyzing verbal and nonverbal cues [3] from spoken interactions can offer insights into behavior-centered domains like psychological assessment and psychotherapy research,

understanding the expression and experience of emotional bond between the interacting client and provider. A previous study [4] found that vocal entrainment, *i.e.*, coordination and adaptation to each other's speech patterns, functions as a marker for empathy in drug addiction counseling. Likewise the nature of dyadic synchrony in vocal arousal between a child and an interacting psychologist was found to be an indicative feature of diagnostic severity of Autism Spectrum Disorder [5]. In this work, we investigate the characteristics of patient-therapist interactions during suicide risk assessment interviews. While researchers have attempted to predict suicide risk itself from speech acoustic features [6, 7], to the best of our knowledge, there has been no work in analyzing suicide risk assessment interviews and interaction dynamics using vocal cues. In addition to the similarity measures used to quantify vocal entrainment [8], the current work employs the notion of complexity in a dynamical systems approach by modeling speech features as nonlinear time-series [9]. For two interlocutors during a conversation, the jointly characterized complexity of their speech patterns can be associated with the degree of their entrainment. More specifically, in case of lower behavioral similarity and less synchrony, there is more variability in the underlying system, resulting into higher system complexity [10].

Equipped with these tools to study spoken interactions of the speakers, we first investigate whether the interaction style during risk assessment interviews is different from other type of interactions. Then we also look into the possibility of using speech-based joint complexity measures as markers of emotional bond between the therapist and the patient.

2. Dataset: Suicide Risk Assessment Corpus

The dataset employed in this paper was collected as a part of research on suicide risk among military personnel [11]. Although 97 active duty Soldiers participated in this study, we currently have complete speech recordings for 61 subjects and annotations for 54 subjects (45 male and 9 female). Based on self-reports, the majority of the participants (75.9%) were Caucasian. The other patients identified themselves as African-American, Native American and Hispanic or Latino. All participants had active suicide ideation during the week preceding the interaction, while 22 of them had attempted suicide at least once in the past. Upon completion of the informed consent process, they were invited for participation in the study.

The study consisted of an interview session followed by crisis intervention and a post-study follow-up on the same day. Five therapists trained in suicide risk assessment and

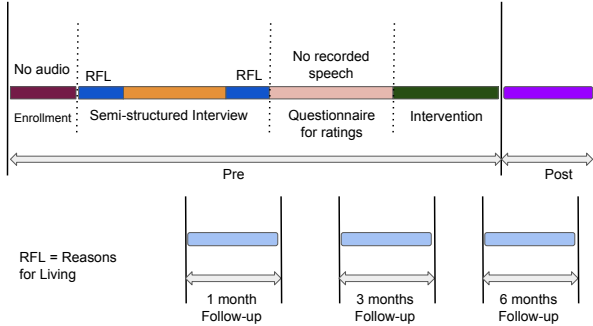


Figure 1: Overview of the structure of a session in the corpus

intervention procedures conducted the study. The conversation took place between the patient and one of the therapists, and was recorded using two directional microphones. The duration of the interview sessions ranged from 10 minutes to 1 hour, varying from patient to patient, resulting into 53 hours in total. The interview session was semi-structured, where the therapist asked a set of questions related to the patient’s reasons for living, elaboration of suicidal thoughts, history of attempts *etc.*. The questions were carefully designed for suicide risk assessment based on the *Beck Scale for Suicidal Ideation (BSSI)* [12] and the *Suicide Attempt Self-Injury Interview (SASII)* structure [13]. Examples of these questions include: “Describe exactly what method you used to injure yourself.”, “With all this going on, what would you say are your reasons for living, or your reasons for not killing yourself?”. The patient’s answer to the latter question, also known as *reasons for living (RFL)* [14], was manually transcribed. After the interview session, an intervention session and a post-study conversation were conducted. The basic structure of a session is shown in Figure 1.

The patient was asked to rate a number of measures: emotional bond, the extent of reasons for living and 10 other attributes related to the patient’s moods (urge, happiness, burden, hope *etc.*) These ratings were self-reported through a written questionnaire and on a scale from 1 to 100. In this work, we only focus on one attribute—emotional bond between the therapist and the patient, as perceived by the patient [15]. For most of the subjects, there were one or more additional follow-up sessions (after 1, 3 and 6 months) which did not have any annotations or transcripts.

3. Feature Extraction

3.1. Preprocessing of the Audio

The audio data of our therapist-patient conversations requires preprocessing in order to extract the speech features in segments spoken by each speaker. First we process the entire audio with voice activity detection (VAD) to identify and remove non-speech or silence regions. For this purpose, we use an RNN-LSTM model-based VAD [16] method that is a part of the OpenSMILE toolkit [17]. Next, we perform speaker diarization to separate the patient and the therapist’s speech. The diarization algorithm [18] used in this work consists of two steps. First, we identify speaker-homogeneous speech segments by detecting the speaker changes using a Generalized Likelihood Ratio (GLR) based criteria. Then we cluster those segments into two separate groups—one for each speaker.

3.2. Frame-level prosodic cues

For analyzing complexity in this work, we primarily rely on two prosodic cues shown to be relevant in psychology and behavioral signal processing literature [19] [3], namely, pitch and energy. We extract speech features with a 25 ms moving window with a 10 ms shift. We use the Praat toolbox [20] to extract pitch and energy. Since pitch extraction is still not very robust and often prone to errors in presence of noise, we perform median filter-based smoothing (window size of 5 samples) on the raw pitch feature stream. In addition, we linearly interpolate the pitch values where pitch is detected to be zero, *e.g.*, unvoiced regions. For energy, we do not perform any postprocessing. Finally, we perform normalization of the feature streams for each speaker.

To compute an alternate PCA-based speech entrainment measure described in Section 5, we extract a 130 dimensional set of features following the proposal of the INTERSPEECH 2013 computational paralinguistics challenge (ComParE) [21], consisting of various prosodic, spectral and voice quality features.

4. Nonlinear Dynamical Systems Modeling of Speech Features

Nonlinear dynamical systems modeling of speech has been used to capture the nonlinearity of vocal characteristics [22] [23]. Such nonlinear dynamical features and complexity measures have shown promising results in many speech applications including speech synthesis [22], phoneme classification [24], speech identification [25] and in the prediction of Parkinson’s disease severity from speech [26]. In our previous work on couples therapy [9], we proposed five complexity measures derived from prosodic features to capture information relevant for studying human behavior.

Using similar techniques from [9], we model the speech feature stream as the observed variable of a nonlinear dynamical system (in other words, a nonlinear time series). We can obtain a state space embedding from the feature stream. Then we compute the largest Lyapunov coefficient and correlation dimension from the state space embedding to quantify the complexity of the underlying dynamical system. In this section we briefly discuss the methods of computing these two complexity measures from speech features.

4.1. State Space Embedding Reconstruction

Let us consider a discrete time-series $z[n]$, *e.g.*, the speech feature stream in our case. Based on the nonlinear dynamical systems approach, the obvious step is to obtain the state space to characterize the underlying dynamics. Assuming a finite-dimensional state space, we are interested in the mapping function Φ that describes the dynamics of the state space: Φ as $\mathbf{x}[n] = \Phi^n(\mathbf{x}[0])$. Unfortunately, finding the original state space from a real-world time-series is an extremely hard problem. However, Takens’ theorem [27] enables us to alternatively construct a mapping from the original state space representation ($\mathbf{x}[n]$) to a reconstructed state space in embedding ($\mathbf{y}[n] \in \mathbb{R}^d$), given by:

$$\mathbf{x} \mapsto \mathbf{y} = (z[n], z[n + \Delta], \dots, z[n + (d - 1)\Delta]) \quad (1)$$

Also known as the delay coordinates mapping, this transformation depends on two important parameters: the embedding dimension, d and the delay Δ . To compute the optimal delay to find the best representation, we consider the

mutual information function of the original time-series and its delayed versions. The location of first local minimum is used as an estimate of Δ . On the other hand, the embedding dimension is estimated using Cao's algorithm [28]. Essentially, this mapping generates a trajectory in d -dimensional state space from the original feature stream by introducing certain delays. Once this trajectory in the reconstructed state space is obtained, analyzing its temporal evolution can provide us with different measures of complexity.

4.2. Different Complexity Measures

4.2.1. Largest Lyapunov exponent (LLE)

A common characterization of complexity in a state space model is its sensitivity to initial conditions, which can be described using the Lyapunov exponents (LE). The Lyapunov exponent of a given direction is defined as the exponential convergence or divergence rate of neighboring trajectories in the state space embedding in that direction. If $\|\delta\mathbf{y}[0]\|$ and $\|\delta\mathbf{y}[n]\|$ denote the Euclidean separation of two trajectories at initial condition and after time n , respectively, and λ_i is the Lyapunov exponent in i^{th} direction,

$$\frac{\|\delta\mathbf{y}[n]\|}{\|\delta\mathbf{y}[0]\|} = e^{\lambda_i n} \quad (n \rightarrow \infty) \quad (2)$$

The largest Lyapunov exponent λ_m can be computed reliably and hence is a popular measure of the complexity. We use an algorithm proposed by Sato *et al.* [29] for estimation of λ_m from the reconstructed state space embedding. A higher value of the largest Lyapunov exponent refers to an increasingly chaotic nature of the system [30].

4.2.2. Correlation Dimension (CD)

Proposed by Grassberger *et al.* [31], the Correlation Dimension is defined as the exponential rate by which a quantity known as correlation sum of points on an embedding grows as a function of radius r . The correlation sum $C(r)$ is defined as the fraction of neighboring points closer than r , averaged over all N points. Mathematically,

$$C(r) \propto r^D \quad (3a)$$

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N \Theta(r - \|\mathbf{y}^i - \mathbf{y}^j\|) \quad (3b)$$

In this equation, $\Theta(\cdot)$ is the Heaviside step function, defined as $\Theta(x) = 1$ for $x \geq 0$ and zero elsewhere. \mathbf{y}^i and \mathbf{y}^j are any two points in the reconstructed state space and D is its correlation dimension. We use a maximum likelihood estimator proposed by Takens [27] to estimate correlation dimension. Intuitively, it provides us with a measure of dimensionality of the reconstructed state space embedding itself.

4.3. Joint Complexity of Therapist-Patient Interaction Dynamics

In this work, we are interested in the style of interaction between the therapist and the patient. Analyzing the complexity of each individual's speech can provide us with characteristics of the speaker, but it does not capture the extent of entrainment. If we can model their interaction with a single nonlinear dynamical system, the similarity or dissimilarity is reflected in the characteristics of that system. This motivates us to consider the speech segments of both speakers to form a combined time-series of features. The speaker normalization step in preprocessing attempts to minimize discontinuities in speaker

boundaries. Complexity measures of the joint system formed by the combined time-series are then computed. However, they also reflect the complexity present within speakers, which is undesirable as indicators of entrainment. Therefore we also compute the complexity of both speakers (therapist and patient) separately and normalize the joint complexity using them, as shown in the following equation:

$$\mathbb{C}(\text{therapist, patient})_{\text{normalized}} = \frac{\mathbb{C}(\text{therapist, patient})}{\sqrt{\mathbb{C}(\text{therapist}) \cdot \mathbb{C}(\text{patient})}}$$

where \mathbb{C} denotes any complexity measure as a function of the speaker (or the dyad), which can be either the largest Lyapunov exponent (LLE) or correlation dimension (CD) in our work. Note that we empirically used the geometric mean of the individual complexity measures of the therapist and the patient for normalization.

5. PCA-based Acoustic Similarity

In addition to complexity measures, we also use a principal component analysis (PCA)-based acoustic similarity metric to quantify vocal entrainment proposed by Lee *et al.* [8]. This method first obtains principal components of features of consecutive speaker turns separately: $Y_1 = X_1^T W_1$ and $Y_2 = X_2^T W_2$. X_1 and X_2 are matrices formed by concatenating feature vectors of all frames in a turn, where each row represents the feature vector of one frame. W_1 and W_2 denote the PCA transformation matrices and the projected feature matrices are Y_1 and Y_2 . If k_1 and k_2 components explain at least 95% of the variance in Y_1 and Y_2 respectively, $k = \max(k_1, k_2)$ components are retained from each PCA representation. Then we form two vectors \mathbf{v}_1 and \mathbf{v}_2 containing the variances of retained components of Y_1 and Y_2 . We obtain two numerically valid probability mass functions p_{11} and p_{22} by normalizing \mathbf{v}_1 and \mathbf{v}_2 , *i.e.*, dividing by their individual vector sum, as in [8]. Similarly, we obtain two other distributions p_{12} and p_{21} by computing variances of $X_1^T W_2$ and $X_2^T W_1$ followed by normalization. Finally, the symmetric Kullback-Leibler (KL) divergence is computed within the pairs (p_{11}, p_{21}) and (p_{22}, p_{12}) and their mean is obtained as the acoustic similarity between the turns:

$$\text{sim}(X_1, X_2) = \frac{1}{2} (D_{sKL}(p_{11} || p_{21}) + D_{sKL}(p_{22} || p_{12})) \quad (4)$$

where $D_{sKL}(p || q)$ is the symmetric KL divergence between p and q . Essentially, it relies on a *bag-of-frames* approach for feature vectors within a speaker turn and computes a similarity metric between consecutive turns of the interlocutors. In order to compute a session-level acoustic similarity measure, we take the mean similarity for all consecutive turns throughout the session.

6. Experiments

6.1. Complexity during Interview Sessions

The risk assessment interview sessions were conducted with predesigned questions. Unlike intervention sessions where the therapist consciously tries to sympathize with the patient to help her or him cope with the crisis, the purpose of interview sessions is to quickly obtain relevant information from the patient. With this objective at hand, the therapist-patient entrainment may not be at par with their entrainment during intervention or follow-up sessions. To test this hypothesis, we conduct an

experiment by computing the normalized joint complexity of therapist-patient pair for the interview, as described in Section 4.3. Each complexity measure (LLE and CD) is used for each of the feature streams (pitch and energy). In addition, PCA-based acoustic similarity measures are computed. We repeat the same operation on the intervention and follow-up sessions and compute the the average complexity of those sessions of the same patient-therapist pair, which we refer to as the baseline complexity. Then we check for what percentage of the subjects the complexity in interview is higher than the baseline complexity. The results are shown in Table 1, where we also present the p -values obtained in the Student’s t-test against the null hypothesis that there is no significant difference in the two aforementioned complexity measures.

Results indicate that the majority (up to about 74%) of the subjects have higher complexity in interviews than their baseline complexity. This observation also turns out to be statistically significant as $p < 0.05$ for all measures. Figure 2 shows the difference in the two complexities for correlation dimension (CD) for energy feature stream by sorting them in increasing order.

Measure	C(interview) > C(baseline)	
	%-age of subjects	p -value [†]
PCA-based similarity*	67.21	0.0072
LLE with pitch	72.13	0.0005
LLE with energy	70.49	0.0014
CD with pitch	65.57	0.0150
CD with energy	73.77	0.0002

Table 1: Results of testing for higher complexity (lower similarity) in interview sessions in comparison to other sessions, i.e., $C(\text{interview}) > C(\text{baseline})$

[†] $p < 0.05$ indicates statistically significant difference
 *we test for similarity(interview) < similarity(baseline) in this case

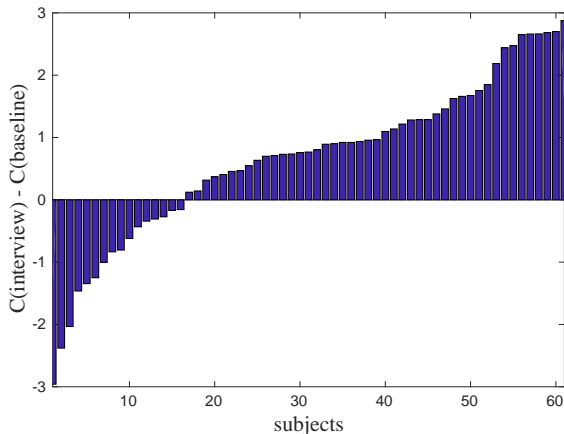


Figure 2: Sorted difference in normalized therapist-patient complexity measure, using correlation dimension (CD) for energy

6.2. Correlation with Emotional Bond

The Pearson’s correlation coefficients between complexity (and similarity) measures are presented in Table 2. All complexity measures are negatively correlated with the emotional bond perceived by the patient ($p < 0.05$) as reported in their survey. The negative values of the correlation coefficients ρ indicate

that higher complexity or lower entrainment is associated with lower emotional bond. Only the PCA-based similarity does not show significant correlation. The reason for this might be the limitation of the PCA-based approach due to inability of capture the temporal dynamics of features within a speaker turn. However, the positive sign of ρ for similarity is consistent with the findings for complexity measures.

Measure	Pearson’s correlation	
	ρ	p -value [†]
PCA-based similarity	0.2480	0.1132
LLE with pitch	-0.3022	0.0419
LLE with energy	-0.3737	0.0148
CD with pitch	-0.2733	0.0473
CD with energy	-0.3815	0.0127

Table 2: Correlation between emotional bond and various complexity (or similarity) measures

[†] $p < 0.05$ indicates statistically significant (strong) correlation

7. Conclusions

In this paper we study complexity measures in speech features to analyze certain mechanisms of dyadic interaction dynamics. We find that joint complexity measures tend to be higher during risk assessment interview sessions of suicide prevention therapy, when compared to the baseline complexity. This indicates a lower degree of therapist-patient entrainment during the interview sessions. Based on the interactions, the patients evaluate their perceived emotional bond with the therapist. We investigate the statistical relationship between ratings of emotional bond and the computed complexity measures. Results show that, joint complexity of speakers, also an opposite notion of entrainment, is negatively correlated with emotional bond. This finding is intuitively justified and consistent with previous studies in psychology [32]. The speech-based approach for analysis of interactions presented in this work can be useful for guidance in conducting more effective interviews in suicide prevention.

This work also offers a number of future directions. In the nonlinear dynamical systems framework, we intend to develop other measures of joint complexity, especially a measure that can capture asymmetry in the interaction dynamics. Given the asymmetric roles of the therapist and patient, such measures could be highly useful. Moreover, analyzing complexity-based measures may reflect characteristics of the patient, particularly his or her ability to connect to the other person during a conversation. A careful analysis of this might be informative towards assessment of the suicide risk itself.

8. Acknowledgements

The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick MD 21702- 5014 is the awarding and administering acquisition office. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Military Suicide Research Consortium under Award No. W81XWH-10-2-0181, and through the Psychological Health and Traumatic Brain Injury Research Program under Award No. W81XWH-15-1-0632. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

9. References

- [1] C. J. Bryan and M. D. Rudd, "Advances in the assessment of suicide risk," *Journal of clinical psychology*, vol. 62, no. 2, pp. 185–200, 2006.
- [2] L. S. Greenberg, J. C. Watson, R. Elliot, and A. C. Bohart, "Empathy," *Psychotherapy: Theory, research, practice, training*, vol. 38, no. 4, p. 380, 2001.
- [3] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE Institute of Electrical and Electronics Engineers*, vol. 101, no. 5, p. 1203, 2013.
- [4] B. Xiao, P. G. Georgiou, Z. E. Imel, D. C. Atkins, and S. Narayanan, "Modeling therapist empathy and vocal entrainment in drug addiction counseling," in *INTERSPEECH*, 2013, pp. 2861–2865.
- [5] D. Bone, C.-C. Lee, A. Potamianos, and S. S. Narayanan, "An investigation of vocal arousal dynamics in child-psychologist interactions using synchrony measures and a conversation-based model," in *INTERSPEECH*, 2014, pp. 218–222.
- [6] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.
- [7] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and D. M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 7, pp. 829–837, 2000.
- [8] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions," *Computer Speech & Language*, vol. 28, no. 2, pp. 518–539, 2014.
- [9] M. Nasir, B. Baucom, S. S. Narayanan, and P. Georgiou, "Complexity in prosody: A nonlinear dynamical systems approach for dyadic conversations; behavior and outcomes in couples therapy," *Interspeech 2016*, pp. 893–897, 2016.
- [10] D. C. Richardson, R. Dale, and K. Shockley, "Synchrony and swing in conversation: coordination, temporal dynamics, and communication," *Embodied communication*, pp. 75–94, 2008.
- [11] C. J. Bryan, J. Mintz, T. A. Clemans, B. Leeson, T. S. Burch, S. R. Williams, E. Maney, and M. D. Rudd, "Effect of crisis response planning vs. contracts for safety on suicide risk in us army soldiers: A randomized clinical trial," *Journal of Affective Disorders*, vol. 212, pp. 64–72, 2017.
- [12] A. T. Beck, M. Kovacs, and A. Weissman, "Assessment of suicidal intention: the scale for suicide ideation," *Journal of consulting and clinical psychology*, vol. 47, no. 2, p. 343, 1979.
- [13] M. M. Linehan, K. A. Comtois, M. Z. Brown, H. L. Heard, and A. Wagner, "Suicide Attempt Self-Injury Interview (SASII): development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury," *Psychological assessment*, vol. 18, no. 3, p. 303, 2006.
- [14] M. M. Linehan, J. L. Goodstein, S. L. Nielsen, and J. A. Chiles, "Reasons for staying alive when you are thinking of killing yourself: the reasons for living inventory," *Journal of consulting and clinical psychology*, vol. 51, no. 2, p. 276, 1983.
- [15] R. L. Hatcher and J. A. Gillaspay, "Development and validation of a revised short version of the working alliance inventory," *Psychotherapy Research*, vol. 16, no. 1, pp. 12–25, 2006.
- [16] F. Eyben, F. Weninger, S. Squartini, and B. Schuller, "Real-life voice activity detection with lstm recurrent neural networks and an application to hollywood movies," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 483–487.
- [17] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the international conference on Multimedia*, 2010, pp. 1459–1462.
- [18] K. J. Han, S. Kim, and S. S. Narayanan, "Strategies to improve the robustness of agglomerative hierarchical clustering under data source variation for speaker diarization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 1590–1601, 2008.
- [19] P. N. Juslin and K. R. Scherer, *Vocal expression of affect*. Oxford University Press, 2005.
- [20] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9/10, pp. 341–345, 2001.
- [21] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi *et al.*, "The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism," 2013.
- [22] M. Banbrook, S. McLaughlin, and I. Mann, "Speech characterization and synthesis by nonlinear methods," *Speech and Audio Processing, IEEE Transactions on*, vol. 7, no. 1, pp. 1–17, 1999.
- [23] S. S. Narayanan and A. A. Alwan, "A nonlinear dynamical systems analysis of fricative consonants," *The Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2511–2524, 1995.
- [24] I. Kokkinos and P. Maragos, "Nonlinear speech analysis using models for chaotic systems," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 6, pp. 1098–1109, 2005.
- [25] A. Petry and D. A. C. Barone, "Speaker identification using nonlinear dynamical features," *Chaos, Solitons & Fractals*, vol. 13, no. 2, pp. 221–231, 2002.
- [26] J. Kim, M. Nasir, R. Gupta, M. V. Segbroeck, D. Bone, M. Black, Z. I. Skordilis, Z. Yang, P. Georgiou, and S. Narayanan, "Automatic estimation of parkinsons disease severity from diverse speech tasks," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
- [27] F. Takens, *Detecting strange attractors in turbulence*. Springer, 1981.
- [28] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1, pp. 43–50, 1997.
- [29] S. Sato, M. Sano, and Y. Sawada, "Practical methods of measuring the generalized dimension and the largest Lyapunov exponent in high dimensional chaotic systems," *Progress of Theoretical Physics*, vol. 77, no. 1, pp. 1–5, 1987.
- [30] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov exponents from time series," *Physical Review A*, vol. 34, no. 6, p. 4971, 1986.
- [31] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D Nonlinear Phenomena*, vol. 9, pp. 189–208, 1983.
- [32] B. Baucom, A. Crenshaw, C. Bryan, T. Clemans, T. Bruce, and M. Rudd, "Patient and clinician vocally encoded emotional arousal as predictors of response to brief interventions for suicidality," *Brief Cognitive Behavioral Interventions to Reduce Suicide Attempts in Military Personnel. Association for Behavioral and Cognitive Therapies*, 2014.