Modeling Therapist Empathy through Prosody in Drug Addiction Counseling

Bo Xiao[1], Daniel Bone[1], Maarten Van Segbroeck[1], Zac E. Imel[2], David C. Atkins[3], Panayiotis G. Georgiou[1], Shrikanth S. Narayanan[1]

[1]SAIL, Dept. Electrical Engineering, University of Southern California, U.S.A.

[2]Dept. Educational Psychology, University of Utah, U.S.A. [3]Dept. Psychiatry & Behavioral Sciences, University of Washington, U.S.A.

[1]http://sail.usc.edu [2]zac.imel@utah.edu [3]datkins@u.washington.edu

This work is supported by NSF, DoD and NIH.

Computational approaches in psychotherapy

Quantitative measures

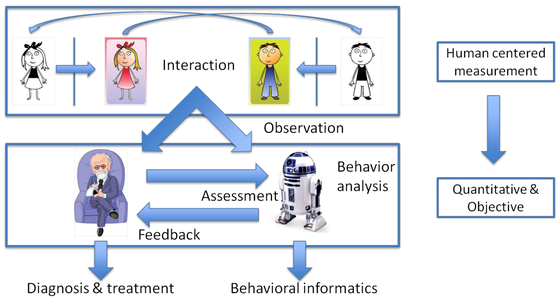
- Illuminate underlying mechanisms of treatment process
- Understand efficacy of psychotherapy treatment approach
- Predict counseling outcome through interaction behavior cues

Empathy

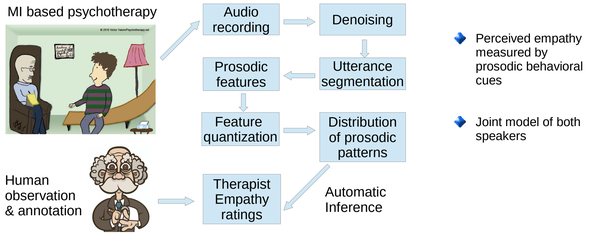
- Feeling for and taking the perspective of others
- Psychological process evident in humans and animals
- Key performance index in addiction counseling
- Associated with positive outcome of interactions

Behavioral Signal Processing

- Human centered approach to modeling human behavior



Proposed empathy modeling framework



Case study dataset

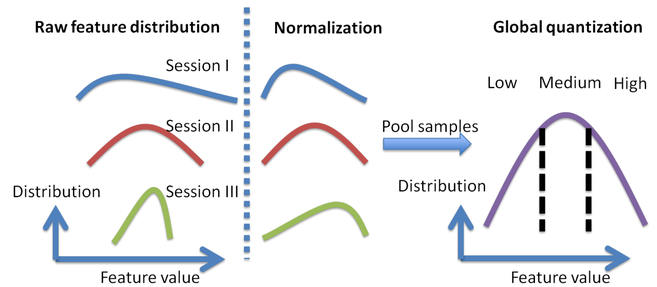
- Counselor training study of Motivational Interviewing (MI)
- MI: emphasize intrinsic motivation of changing addiction
- Three coders; score range 1 to 7; 836 sessions in total
- Selected 71 high & 46 low scored sessions, 20 min each

Prosodic features

Prosody

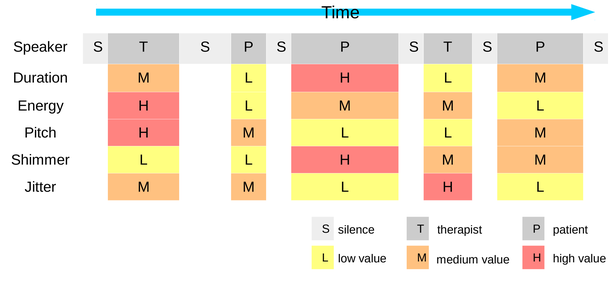
- Intonation and rhythmic aspects of speech (*how* one says)
- Neurological and empirical evidence of relation to empathy
- Prosodic features: energy, pitch, shimmer, jitter, duration

Proposed representation



- Compute averaged prosodic features for each utterance
- Normalize features per speaker per session
- Find equal-size quantization points of features in training set
- Reduce dimensionality, results become more interpretable

Distribution of prosodic patterns



- Session level joint distribution of quantized prosodic features

$$f_n^i \in \{\texttt{Duration}, \texttt{pitch}, \texttt{jitter}, \texttt{energy}, \texttt{shimmer}\},\ n = 1, 2, \cdots, N$$

- Prosodic pattern distribution of a single utterance

$$P_U(r_n, F_n),\ r_n \in \{\texttt{T}, \texttt{P}\},\ F_n \in \{(f_n^1), (f_n^1, f_n^2), (f_n^1, f_n^2, f_n^3)\}$$

- Prosodic pattern distribution of neighboring utterances

$$P_U(r_n, F_n, r_{n+1}, F_{n+1}), r_n \in \{\texttt{T}, \texttt{P}\}, F_n \in \{(f_n^1)\}$$

Results

Correlation analysis — Prosody & Empathy

- High pitch and high energy point to low empathy
- Salient pattern (T, $d = M, p = H, e = H$): 6% of T's utt.

Speaker	Prosodic feature patterns			Corr.
T	M duration	H pitch	H energy	-0.47
T	M duration	H pitch		-0.42
T	M duration	H energy	M shimmer	-0.41

Speaker 1	Feature 1	Speaker 2	Feature 2	Corr.
T	M energy	T	M energy	-0.40
T	M jitter	T	H jitter	-0.34
P	H duration	T	L duration	0.34

In total 51 patterns correlated (p<0.001) $|\rho| > 0.3$

Conditioned on therapist utterances

- Low energy positive, high pitch/energy still negative

Prosodic feature patterns			Corr.
M duration	H pitch	H energy	-0.33
L duration	L energy	H shimmer	0.31
L energy			0.30

Classification of high/low empathy by prosody

- Leave-one-therapist-out cross validation (total 91) by SVM

Approach	Accuracy
Chance level	61%
Functionals of prosodic features	67%
Vocal similarity and turn taking ratio	70%
Distribution of prosodic patterns P_U	75%

Findings & Contributions

- Prosodic correlates of perceived therapist empathy
- Quantization and joint modeling of prosody derives salient/indicative prosodic patterns
- In the future:
  - Joint modeling of prosody and lexical information
  - Larger scale experiments and clinical translation