Understanding and Predicting Empathic Behavior in Counseling Therapy

Verónica Pérez-Rosas[1], Rada Mihalcea[1], Kenneth Resnicow[2]
Satinder Singh[1] and Lawrence An[3]
[1]Computer Science and Engineering, [2]School of Public Health
[3]Center for Health Communications Research
University of Michigan
{vrncapr,mihalcea,kresnic,baveja,lcan}@umich.edu

Abstract

Counselor empathy is associated with better outcomes in psychology and behavioral counseling. In this paper, we explore several aspects pertaining to counseling interaction dynamics and their relation to counselor empathy during motivational interviewing encounters. Particularly, we analyze aspects such as participants' engagement, participants' verbal and nonverbal accommodation, as well as topics being discussed during the conversation, with the final goal of identifying linguistic and acoustic markers of counselor empathy. We also show how we can use these findings alongside other raw linguistic and acoustic features to build accurate counselor empathy classifiers with accuracies of up to 80%.

1 Introduction

Behavioral counseling is an important tool to address public health issues such as mental health, substance abuse, and nutrition problems among others. This has motivated increased interest in the study of mechanisms associated with successful interventions. Among them, counselor empathy has been identified as a key intervention component that relates to positive therapy outcomes.

Displaying empathic behavior helps counselors to build rapport with their clients. Empathy levels experienced during counseling have a significant effect on treatment outcomes, as clients who perceive their counselor as empathic are more likely to improve than the ones who do not (Moyers and Miller, 2013).

In this paper, we apply quantitative approaches to understand the dynamics of the counseling interactions and their relation to counselor empathy. We focus our analysis on counseling conducted using Motivational Interviewing (MI), a well-established evidence-based counseling style, where counselor empathy is defined as the active interest and effort to understand the client's perspective (Miller and Rollnick, 2013).

We address four main research questions. First, are there differences in how the counselor and the client engage during empathic conversations? We explore this question by conducting turn-by-turn word frequency analyses of participant's interactions across the counseling conversations. Second, are there differences in verbal and vocal mimicry patterns occurring during high and low empathy interactions? We address this question by measuring the degree of language matching, verbal and nonverbal coordination, and power dynamics expressed during the interaction. Third, are there content differences in counselor discourse during high and low empathy interactions? We answer this question by applying topic modeling to identify the topics that are more salient in high and low empathy interventions (or in both). Finally, fourth, can we build accurate classifiers of counselor empathy? We show how the linguistic and acoustic empathy markers identified in our analyses, together with other raw features, can be used to construct classifiers able to predict counselor empathy with accuracies of up to 80%.

2 Related Work

There have been several efforts to study the role of empathy during counseling interactions. (Xiao et al., 2012) applied a text-based approach to discriminate empathic from non-empathic encounters using word-frequency analysis. They conducted a set of experiments aiming to predict empathy at the utterance and session level on a manually annotated dataset. Results showed that empathy can

be predicted at reasonable accuracy levels, comparable to human assessments. (Gibson et al., 2015) presented a more refined approach for this task, which in addition to n-grams included features derived from the Linguistic Inquire Word Count, LIWC (Tausczik and Pennebaker, 2010) as well as psycholinguistic norms.

Other research has focused on exploring aspects related to counselor empathy skills, such as their ability to match the client language. (Lord et al., 2015) analyzed the language coordination between client and counselor using Language Style Synchrony (LSS), a measure of the degree of similarity in word usage among speakers in adjacent talking turns. They found that empathy scores are positively related to LSS, and that higher levels of LSS are likely to result in higher empathy scores.

Another line of work has explored the use of the acoustic component to predict empathy levels during counseling encounters. (Xiao et al., 2014) presented a study on the automatic evaluation of counselor empathy based on the analysis of correlation between prosody patterns and the degree of empathy showed by the therapist during the counseling interactions. More recently, (Xiao et al., 2015) addressed the empathy prediction task by deriving language models from transcripts obtained by an automatic speech recognition system, thus eliminating the need of human intervention during speaker segmentation and transcription.

Most of this previous research has focused on the prediction task, and explored a variety of linguistic and acoustic representations for this goal. While some of this work has explored the linguistic accommodation between speakers, previous methods have not fully explored the conversational aspects of the counseling interaction.

In this paper, we seek to explore how conversational aspects such as engagement, accommodation, and discourse topics are related to counselor empathy by using strategies such as turn-by-turn word frequency analysis, language coordination, power dynamics analysis, and topic modeling. Furthermore, we build accurate empathy classifiers that rely on acoustic and linguistic cues inspired by our conversational analyses.

3 Counseling Empathy Dataset

The dataset used in this study consists of 276 MI audio-recorded sessions from: two clinical research studies on smoking cessation and medica-

tion adherence (Catley et al., 2012; Goggin et al., 2013); recordings of MI students from a graduate-level MI course; wellness coaching phone calls; brief medical encounters in dental practice and student counseling. The dataset was obtained from a previous study conducted by the authors. Further details can be found in (Pérez-Rosas et al., 2016).

The counseling sessions target three behavior changes: diet changes (72 sessions), smoking cessation (95 sessions), medication adherence (93 sessions). In addition, there are 16 sessions on miscellaneous topics. The full set comprises 97.8 hours of audio with an average session length of 20.8 minutes with a standard deviation of 11.5 minutes.

3.1 Data Preprocessing

Before conducting our analysis on the collected dataset, we performed several preprocessing steps to ensure the confidentiality of the data and to enable automatic text and audio feature extraction.

First, all the counseling recordings were subjected to an anonymization process. This includes manually trimming the audio to remove introductions, and inserting silences to replace references to participant's name and location.

Next, 162 sessions for which transcripts were not readily available were transcribed via Mechanical Turk (Marge et al., 2010) using the following guidelines: 1) transcribe speech turn by turn, 2) clearly identify the speaker (either client or counselor), 3) include speech disfluencies, such as false starts, repetitions of whole words or parts of words, and fillers. Transcriptions were manually verified at random points to avoid spam and ensure their quality.

Since sessions were recorded in natural conditions, we applied speech enhancement methods to remove noise and improve the speech signal quality. We started by converting the audio signal from a stereo to a mono channel and to a uniform sample rate of 16k. We then applied the Mean Square Error estimation of spectral amplitude for audio denoising, as implemented in the Voicebox Speech Processing toolbox (Brookes, 2003). To allow for a turn-by-turn audio analysis of the counseling interaction, we processed the speech signal to separate client and counselor speech segments. To accomplish this task, we used on automatic speech-to-text forced alignment API.[1] We

---

[1]YouTube Data API

then used the automatically-obtained time stamps to segment the audio and derive speaker-specific speech segments for each counseling dyad.

3.2 Data Annotation

Empathy assessments were obtained using the Motivational Interviewing Treatment Integrity (MITI) coding scheme version 4.1 (Moyers, 2014). Each session was assigned an empathy score using a 5-point Likert scale, which measures the extent to which the clinician understands or makes an effort to grasp the client's perspective and feelings. The coding was conducted by two independent teams of three coders who had previous experience in MI and MI coding. Annotations were conducted using the session audio recording along with its transcript. The inter-rater reliability, measured in a random sample of 20 double coded sessions using the Intra-Class Correlation Coefficient was 0.60,[2] suggesting that the annotators showed moderate agreement on empathy assessments. The reported annotation agreement was calculated on the original 5-scale empathy score and it is within the ranges reported in previous Motivational Interviewing studies (0.60-0.62). Because of the skewed frequency distribution of the empathy scores in the dataset, we decided to conduct our analyses using empathy as a binary outcome, by classifying scores from 1 to 3 as low empathy, and scores of 4 and 5 as high empathy. This resulted in 179 high empathy sessions and 97 low empathy sessions.

4 Empathic vs Non-Empathic Interactions: Counselor Engagement

We start by exploring differences in verbal exchange length between low and high empathy encounters as an indirect measure of participants engagement during the conversation. In this analysis, we account for the time dimension by segmenting the conversation into five equal portions. First, we look at the ratio of words exchanged between the counselor and the client for the different fractions of the conversation.[3] As shown in Figure 1, low empathy interactions present noticeably
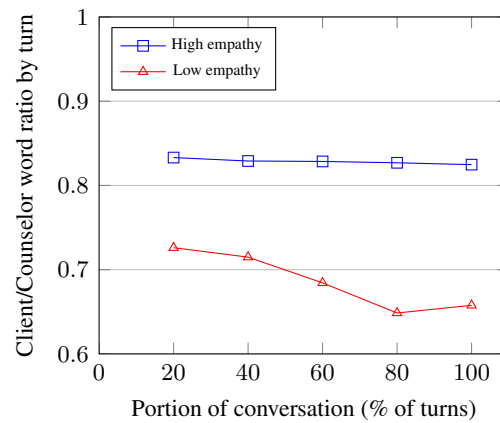

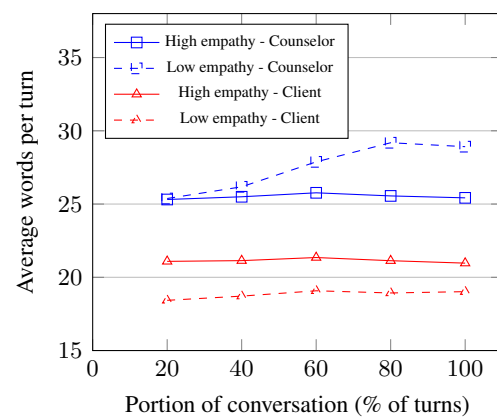Figure 1: Word ratio by turn between clients and counselors as the conversation progresses.

Figure 2: Average words per turn by counselors and clients as the conversation progresses.

lower ratio of words exchanged between counselors and clients across the interaction, while high empathy exchanges show consistently higher levels of interaction. This can be further observed in Figure 2, which shows that more empathic counselors speak considerably less than their clients, and that their less empathic counterparts. This is in line with findings in MI literature indicating that counselors who reduce the amount of time they talk with their clients are likely to allow more time for the patient to talk and explore their concerns, thus improving the perception of empathy and understanding.

5 The Role of Verbal and Nonverbal Accommodation in Empathy

Accommodation in health care communication involves counselor and client coordination including participation in communication and decision making, and shared understanding (D'Agostino

---

[2]ICC scores were obtained using a two-way mixed model with absolute agreement.

[3]This ratio is calculated for each pair of turns in the conversation, and it is simply measured as the number of words uttered by the counselor divided by the number of words uttered by the client. The turn-level word ratios are then averaged for all the turns included in a portion of the conversation.

and Bylund, 2014). We analyze the accommodation and its relation to empathy by exploring verbal and nonverbal behaviors exhibited by counseling participants during MI encounters. In addition to accommodation assessments, we explore the direction of the accommodation phenomena, i.e., whether the counselor is mirroring or leading the client.

5.1 Verbal Accommodation

In order to explore how verbal accommodation phenomena in our dataset relate to the MITI empathy assessments, we use two methods that are drawn from the Conversation Accommodation Theory. The first one is the Linguistic Style Matching (LSM) proposed in (Gonzales et al., 2009) to quantify to which extent one speaker, i.e., the counselor, matches the language of the other, i.e., the client. The second one is the Linguistic Style Coordination (LSC) metric proposed in (Danescu-Niculescu-Mizil et al., 2011), which quantifies the degree to which one individual immediately echoes the linguistic style of the person they are responding to. Both metrics are evaluated across eight linguistic markers from the LIWC dictionary (Tausczik and Pennebaker, 2010) (i.e., quantifiers, conjunctions, adverbs, auxiliary verbs, prepositions, articles, personal pronouns and impersonal pronouns).

LSM produces a score ranging between 0 and 1 indicating how much one person uses types of words comparable to the other person, while LSC generates a coordination score in the range of -1 to 1 indicating the degree of immediate coordination between speakers. While both measures are designed to analyze verbal synchrony, they can reveal different aspects of the counseling interaction. We use LSM to explore the potential match of language between counselors and clients across the counseling interaction, and we use LSC to quantify whether the counselor use of a specific linguistic marker in a given turn increases the probability of the client using the same marker during their reply. In addition, we use LSC to investigate power differences during the conversation based on the amount of coordination displayed by either counselor or client, under the assumption that the speaker who accommodates less holds the most power during the conversation (Danescu-Niculescu-Mizil et al., 2012).
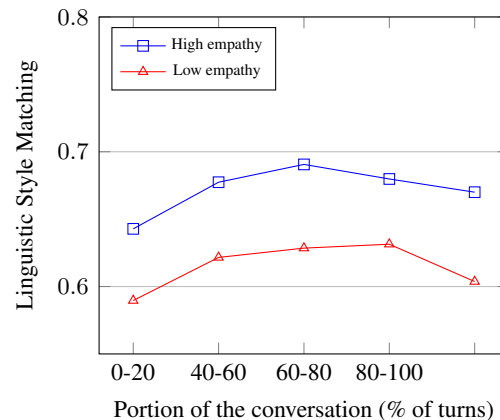
Figure 3 shows the average LSM scores for

Figure 3: Linguistic style matching across five equal segments of the conversation duration.

eight linguistic markers measured on five equal segments of the conversation duration. As expected, we observe an increasing trend of language style matching during the counseling interaction in both high-empathic and low-empathic encounters, as people usually match their language unconsciously and regardless of the outcome of the conversation (Niederhoffer and Pennebaker, 2002). Interestingly, counselors and clients present a higher degree of language matching during high empathy encounters, while speakers in low empathy encounters show lower levels of style matching.

We evaluate the immediate LSC in two directions: coordination of counselors toward clients, and coordination of clients toward counselors. Results indicate low levels of immediate coordination in both cases, with values ranging between -0.06 and 0.1. Nonetheless, the results also suggest that clients coordinate more than counselors, with LSC(client,counselor)=-0.030 compared to LSC(counselor, client)=-0.038, which further suggests that counselors have more power (control) during the conversation.[4]

Analyses of the LSC levels from counselors to clients on different linguistic markers across high-empathic and low-empathic interactions provide interesting findings. While counselors generally show lower levels of coordination in the use of prepositions, auxiliary verbs, and personal pronouns (Figure 4), low-empathic counselors show higher LSC levels than their high-empathic counterparts. This can be attributed to the use of con-

---

[4]The differences in coordination showed during the analyses are statistically significant (two tailed t-test, p=0.0156)
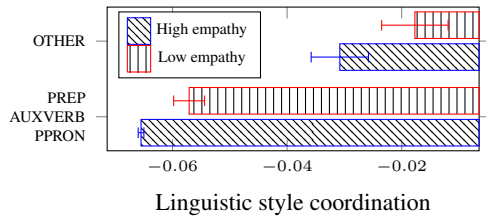
Figure 4: Linguistic style coordination from counselors to clients. OTHER include: quantifiers, conjunctions, adverbs, articles, and impersonal pronouns.
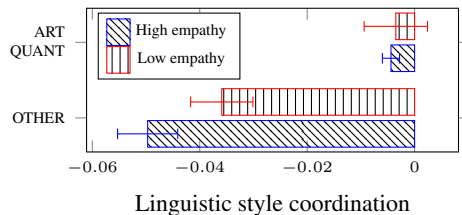

Figure 5: Linguistic style coordination from clients to counselors. OTHER include: conjunctions, adverbs, auxiliary verbs, prepositions, personal pronouns and impersonal pronouns.

frontational language (e.g., *I*, *could*, *should*, and *have*), which is often associated with low empathy. Similar analyses on the client side, shown in Figure 5, indicate significant differences in the use of linguistic markers by the client (except for articles and quantifiers). In particular, during low empathy encounters, clients coordinate more on the use of conjunctions, adverbs, auxiliary verbs, prepositions, personal pronouns, and impersonal pronouns.

5.2 Nonverbal Accommodation

Empathy is also shown through nonverbal channels such as visual and acoustics (Regenbogen et al., 2012). We explore the role of nonverbal mirroring in empathy by looking at vocal synchrony patterns shared between counselors and clients during the counseling interaction. We focus our analysis on vocal pitch, which is defined as the psychological perception of the voice frequency in terms of how high or how low it sounds. Pitch carries information about the speaker's emotional state, and has been shown to be related to the perception of empathy in psychotherapy (Reich et al., 2014).

We evaluate speech synchrony during turn-taking trajectories in the conversation. We con-
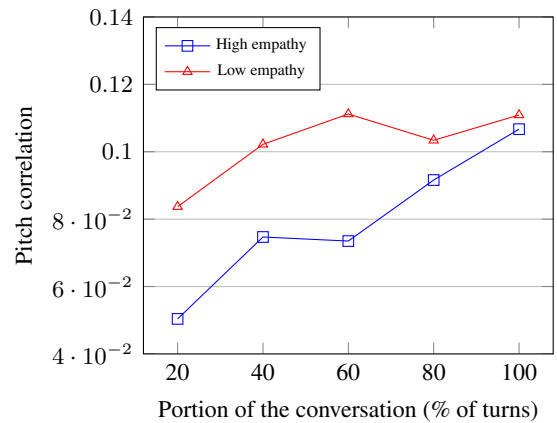

Figure 6: Pitch correlation among participants during counselor following turns as the conversation progresses.

sider two cases: sequences where the counselor replies to the client statements (e.g., rephrasing), and sequences where the counselor leads the interaction (e.g., asking questions). Starting with the turn-by-turn segmentation,[5] we extract pitch (F0) on each speaker-specific segment using OpenEar (Eyben et al., 2009).[6] We then measure the correlation of all pitch values during counselor following turns and during counselor leading turns across the entire therapy session.[7]

Figures 6 and 7 show the trends in pitch synchrony across high-empathic and low-empathic encounters in the dataset. In the first figure, we observe that when replying to clients, counselors who are given low empathy scores show higher vocal synchrony levels than counselors who receive higher empathy scores. A potential explanation for this finding is that a counselor who mirrors the client pitch might amplify the emotional distress of the client, or may suggest the counselor's lack of confidence or knowledge (Reich et al., 2014).

On the other hand, we observe the opposite trend for the counselor leading sequences, where higher vocal synchrony levels are observed during high empathy interactions, which can be at-

---

[5] On average, there are approximately 40 counselor-client turns per conversation

[6] The feature extraction was done at audio-frame level every 10ms with a 25ms Hamming window.

[7] The terms of "counselor following" and "counselor leading" simply refer to how the correlation is computed. In "counselor following," we consider the set of counselor utterances and the previous client utterances; in "counselor leading," we consider the set of counselor utterances and the following client utterances.
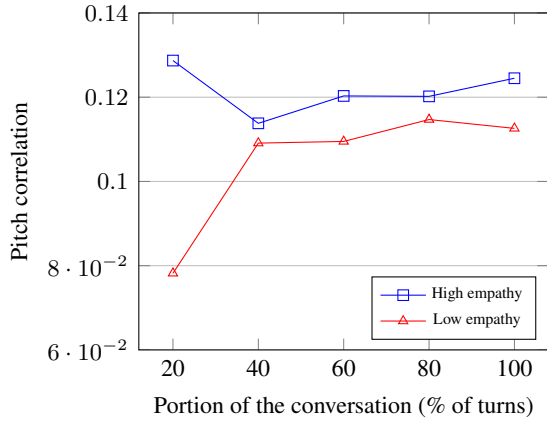
Figure 7: Pitch correlation among participants during counselor leading turns as the conversation progresses.

Behavior target	Sample words
Medication adherence	Adherence, dose, window, target, adherent, maintain, track
Smoking cessation	Cigarette, nicotine, risk, addiction, smoke, withdrawal
Weight management	diet, weight, eat, food, meal, lose, gain, cook, exercise

Table 1: Three behavior change targets in the dataset

tributed to clients mirroring the counselor speech. The similarity is noticeably higher at the beginning of the conversation and gradually decreases as the conversation progresses. Moreover, the differences are not significant for the 40-100% turns, but results for the first 20% suggest significant differences at least in the beginning of the conversation ($p < 0.05$). This further confirm similarities during verbal and nonverbal accommodation, similar to how in section 5.1 we found that during high-empathic encounters, counselors hold control of the conversation and clients accommodate more than counselors.

6 Topics Discussed during Counseling Interaction and their Relation to Empathy

We also conduct content analysis on the counseling interactions, to identify themes discussed in high-empathic and low-empathic encounters. For this task, we employ the Meaning Extraction Method (MEM) (Chung and Pennebaker, 2008), a topic extraction method that identifies the most common words used in a set of documents, and clusters them into coherent themes by analyzing their co-occurrences. MEM has been used in the past in the psychotherapy domain to analyze salient topics in depression forums (Ramirez-Esparza et al., 2008) and also to investigate differences in topics discussed by patients given their therapy outcomes, i.e., therapeutic gain or unsuccessful therapy (Wolf et al., 2010).

Our analyses are conducted on counselor turns only, thus all the client turns are removed from each session transcript. We use the Meaning Extraction Helper tool (Boyd, 2016) to conduct the text preprocessing tasks, which include tokenizing and lemmatizing the words in each session, as well as removing function words. We keep only words who appear in at least 10% of the transcripts with a minimum frequency of 50. From the resulting list, we remove adjectives, adverbs, and verbs and keep only nouns as they usually refer to one definite class thus helping us to identify less ambiguous topics. Using the resulting noun list, with 339 entries, we generate a binary vector for each document, indicating the presence or absence of each noun in the document. We then run a Principal Component Analysis (PCA), followed by varimax rotation on the document matrix to find clusters of co-occurring nouns.

The initial PCA shows that the first three components consist mainly of domain specific nouns. Notably, this accurately captures the presence of the three main behavior change targets discussed in the dataset, i.e., medication adherence, smoking cessation, and weight management; sample words from each component are shown in Table 1.

In order to identify topics potentially related to the counseling skill, we decided to remove the domain words from the analysis, which resulted in 250 nouns. Next, we use the same PCA configuration on the binary document matrix and rerun the experiment, which this time leads to 98 components. Following PCA literature recommendations (Velicer and Fava, 1998), we retain only components with at least three variables with loadings greater than 0.30, which leads to 14 components. We then re-run PCA forcing a 14 components solution; these components explain 35% of the total variance in the original matrix. Finally, we use the method proposed in (Wilson et al., 2016) to measure the degree to which a particular MEM topic (component) is used during high-empathic and low-emphatic encounters.

Topic	Sample nouns	Score
Concerns	Concern, scare, overwhelm, diagnose	2.52
Importance	Importance, reason, maintain, sense, increase	1.41
Inform	Information, schedule, discuss, read	1.14
Reflections	sound, start, look, mention, past, notice	1.19
Change	Health, past, experience, decision, realize, difficult, impact	1.27
Goals	Reach, choose, period, stick, idea, study, record	1.57
Motivation	Plan, motivate, routine, motivation, group, progress, fun	1.10
Support	Family, care, worry, job, lifestyle, job, focus, issue	0.92
Feelings	Worry, deal, stuck, struggle, leave	0.91
Guide	Stop, reduce, attempt, spend, replacement	0.79
Resistance	Trouble, barrier, fear, reach, involve, cover	073
Persuade	Routine, track, strategy, recommend, affect	0.64
Persuade	Stop, increase, decrease, benefit, consequences	0.455
Plan	Activity, strategy, barrier, couple	0.292

Table 2: Topics extracted by the MEM from MI sessions, along with sample nouns and salient topic scores.

Table 2 shows the scores assigned to each topic. In this table, scores greater than 1 correspond to topics salient in high empathy encounters while scores lower than 1 indicate topics salient in low empathy encounters. From this table, we can derive interesting observations. First, during high-empathic encounters, counselors seem to pay more attention to patient concerns, provide information, use reflective language, and talk about change. Second, during less empathic encounters, counselors seem to persuade and direct more, as well as focus on client's resistance to change. Interestingly, topics that are identified as dominant in less empathic interactions are also related to MI non-adherent behavior, which means the counselors are not following the MI strategy (Rollnick et al., 2008). Finally, regardless of the empathy shown during the encounter, counselors discuss patients' support system and feelings at similar rates (values closer to 1), which is expected when following the MI strategy.

7 Prediction of Counselor Empathy

In the previous sections, we provided evidence of important differences in linguistic and verbal behaviors exhibited by counselors and clients during high-empathic and low-empathic MI encounters. In this section, we explore the use of linguistic and acoustic cues to build a computational model that predicts counselor empathy during MI encounters.

The feature set consists of the cues identified during our exploratory analyses as potential indicators of counselor empathy, as well as additional text and audio features used during standard NLP and acoustic feature extraction.

The text-based features are extracted from the manual transcripts of the sessions, while the audio-based features are extracted from audio segments obtained by force-aligning each session transcript with its corresponding audio. However, as future work, we are considering to automatize this process by conducting automatic speaker diarization and transcription via automatic speech recognition.

During our experiments, we first explore the predictive power of each cue separately, followed by an integrated model that attempts to combine the linguistic and acoustic cues to improve the prediction of counselor empathy.

All the experiments are performed using a Random Forest (Breiman, 2001) classifier. Given the large number of features, we use feature selection based on information gain to identify the best set of features during each experiment. During this process we keep at least 20% of the features in each set. Evaluations are conducted using leave-one-session-out cross-validation. The feature selection algorithm is run on each training fold before the model is trained, and the final model includes the best subset of attributes. As a reference value, we use a majority class baseline, obtained by selecting high empathy as the default class, which corresponds to 64% accuracy.

7.1 Linguistic and Acoustic Features

Engagement: These features represent the participant's engagement during the conversation as described in Section 4. They are evaluated at 20% increments of the conversation duration and also at conversation (session) level. The features are listed in Table 3.

Linguistic accommodation: We measure the LSM and LSC metrics as described in section 5.2 over 74 LIWC categories and measured at 20% increments of the encounter duration.

---

[8]Calculated using the LSC metric

Feature	C	T
Counselor talk time based on syllable counting	✓	
Length of conversation setter, length of setter response, ratio between setter and response	✓	
Counselor turns, client turns	✓	✓
Average words during client and counselor turns	✓	✓
Ratio of counselor and client words in each turn	✓	✓
Rate of verbal mirroring on each LIWC category [8]	✓	

Table 3: Engagement features extracted at a) (C) conversation level, and b) (T) 20% increments of the conversation duration, in percentage of turns.

Nonverbal accommodation: This set includes the counselor-leading and counselor-following synchrony scores, calculated as described in section 5.2, and evaluated at 20% increments of the encounter duration.

Discourse topics: These features consist of the 14 topics identified in section 6 as frequently discussed during the MI encounters. The features are obtained by calculating the product of the principal components matrix and the binary document-term matrix.

Raw linguistic features: We extract a large set of linguistic features derived from the session transcript to model the counselor language. We include: unigrams and bigrams (ngrams), represented as a vector containing their frequencies in the session; psycholinguistic-inspired features that capture differences in semantic meaning (lexicons), represented as the total frequency counts of all the words in a lexicon-category that are present in the transcript; syntactic features that encode syntax patterns in the counselor statements (CFG), represented as a vector containing the frequency of lexicalized and unlexicalized production rules from the Context Free Grammar parse trees[9] of each transcript. The final linguistic features set consists of 13,648 features.

Raw acoustic features: This feature set includes a large number of speech features extracted with the OpenEar toolkit (Eyben et al., 2009). We use a predefined feature set, EmoLarge, which consists of a set of 6,552 features used for emotion recognition tasks. The features are derived from 25 low-level speech descriptors including intensity, loudness, 12 Mel frequency coefficients, pitch (F0),

---

[9]Extracted with the Stanford parser (Klein and Manning, 2003).

Feature set	Empathy		
		F-score	
	Acc.	HE	LE
Linguistic			
Engagement	71.01%	0.80	0.40
Ling Accom	73.19%	0.82	0.44
Topics	75.72%	0.83	0.57
N-grams	78.62%	0.86	0.58
Lexicons	76.09%	0.84	0.55
CFG	76.09%	0.84	0.53
All linguistic	**80.07%**	**0.87**	0.62
Acoustic			
Nonverb Accom	64.86%	0.79	0.00
Raw acoustic	73.91%	0.82	0.53
All acoustic	75.72%	0.83	0.56
Linguistic+ Acoustic			
Ling+acoustic (early)	76.81%	0.84	0.56
Ling+acoustic (late)	79.35%	0.86	**0.71**

Table 4: Overall prediction results and F-scores for high empathy (HE) and low empathy (LE) using linguistic and acoustic feature sets.

probability of voicing, F0 envelope, zero-crossing rate, and 8 line spectral frequencies.

7.2 Classification Results

Classification results for each feature set are shown in Table 4. For the linguistic and acoustic modalities, almost all the feature sets provide classification accuracies above the baseline, with good F-scores for both high and low empathy. The only exception are the nonverbal accommodation features, which have an accuracy comparable to the baseline (64.86% vs. 64%).

When combining all the feature sets for each modality, we observe performance gains in the range of 10 to 15%, as compared to the models that use one feature set at a time.

We also conduct multimodal experiments where we combine linguistic and acoustic features using either early fusion by concatenating all the feature vectors, or late fusion by aggregating the outputs of each classifier using a rule-based score level fusion that assigns a weight of 0.8 to the linguistic classifier, and 0.2 to the acoustic classifier.[10]

Overall, the results show performance gains when using late fusion as compared to early fusion. While the late fusion model does not outperform the best linguistic model in terms of accuracy and high empathy F-score, the multimodal late fusion classifier has significantly better F-score performance in the classification of low empathy encounters, thus suggesting potential benefits of fus-

---

[10]Weights empirically determined on a development set by evaluating increments of 0.2 for each classifier weight.

ing acoustic and linguistic cues during the prediction of counselor empathy.

8 Conclusions

In this paper, we presented an extensive analysis of counselors and clients behaviors during MI encounters, and found significant differences in the way counselors and clients behave during high and low empathy encounters. We specifically explored the engagement, coordination, and discourse of counselors during MI interventions. Our main findings include:

Engagement: Empathic counselors show more engagement during the conversation by a) showing levels of verbal interaction consistent with their client, and b) reducing their relative talking time with clients.

Coordination: Empathic counselors match the linguistic style of their clients across the session, but maintain control of the conversation by coordinating less at immediate conversation turn-level.

Conversation content: Empathic counselors use reflective language and talk about behavior change, while less empathic counselors persuade more and focus on client resistance toward change.

The results of these analyses were used to build accurate counselor empathy classifiers that rely on linguistic and acoustic cues, with accuracies of up to 80%.

In the future, we plan to build upon the acquired knowledge and the developed classifiers to create automatic tools that provide accurate evaluative feedback of counseling practice.

Acknowledgments

References

Ryan. L. Boyd. 2016. Meh: Meaning extraction helper (version 1.4.14) [software].

Leo Breiman. 2001. Random forests. *Machine learning* 45(1):5–32.

Michael Brookes. 2003. Voicebox: Speech processing toolbox for matlab.

Delwyn Catley, Kari J Harris, Kathy Goggin, Kimber Richter, Karen Williams, Christi Patten, Ken Resnicow, Edward Ellerbeck, Andrea Bradley-Ewing, Domonique Malomo, et al. 2012. Motivational interviewing for encouraging quit attempts among unmotivated smokers: study protocol of a randomized, controlled, efficacy trial. *BMC public health* 12(1):456.

Cindy K Chung and James W Pennebaker. 2008. Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality* 42(1):96–132.

Thomas A D'Agostino and Carma L Bylund. 2014. Nonverbal accommodation in health care communication. *Health communication* 29(6):563–573.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World Wide Web*. ACM, pages 745–754.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of WWW*. pages 699–708.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2009. Openear introducing the munich open-source emotion and affect recognition toolkit. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, pages 1–6.

James Gibson, Nikolaos Malandrakis, Francisco Romero, David C Atkins, and Shrikanth Narayanan. 2015. Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Kathy Goggin, Mary M Gerkovich, Karen B Williams, Julie W Banderas, Delwyn Catley, Jannette Berkley-Patton, Glenn J Wagner, James Stanford, Sally Neville, Vinutha K Kumar, et al. 2013. A randomized controlled trial examining the efficacy of motivational counseling with observed therapy for antiretroviral therapy adherence. *AIDS and Behavior* 17(6):1992–2001.

Amy L Gonzales, Jeffrey T Hancock, and James W Pennebaker. 2009. Language style matching as a predictor of social dynamics in small groups. *Communication Research* .

Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, pages 423–430.

Sarah Peregrine Lord, Elisa Sheng, Zac E Imel, John Baer, and David C Atkins. 2015. More than reflections: Empathy in motivational interviewing includes language style synchrony between therapist and client. *Behavior therapy* 46(3):296–303.

Matthew Marge, Satanjeev Banerjee, Alexander Rudnicky, et al. 2010. Using the amazon mechanical turk for transcription of spoken language. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, pages 5270–5273.

William R Miller and Stephen Rollnick. 2013. *Motivational interviewing: Helping people change, Third edition*. The Guilford Press.

Theresa B. Manuel Jennifer K. Ernst Denise Moyers. 2014. *Motivational Interviewing Treatment Integrity Coding Manual 4.1. Unpublished manual.*.

Theresa B Moyers and William R Miller. 2013. Is low therapist empathy toxic? *Psychology of Addictive Behaviors* 27(3):878.

Kate G Niederhoffer and James W Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21(4):337–360.

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Lawrence An, and Satinder Singh. 2016. Building a motivational interviewing dataset. In *NAACL Workshop on Clinical Psychology*.

Nairan Ramirez-Esparza, Cindy K. Chung, Ewa Kacewicz, and James W. Pennebaker. 2008. The psychology of word use in depression forums in english and in spanish: Testing two text analytic approaches. In *In Proc. ICWSM 2008*.

Christina Regenbogen, Daniel A Schneider, Andreas Finkelmeyer, Nils Kohn, Birgit Derntl, Thilo Kellermann, Raquel E Gur, Frank Schneider, and Ute Habel. 2012. The differential contribution of facial expressions, prosody, and speech content to empathy. *Cognition & emotion* 26(6):995–1014.

Catherine M Reich, Jeffrey S Berman, Rick Dale, and Heidi M Levitt. 2014. Vocal synchrony in psychotherapy. *Journal of Social and Clinical Psychology* 33(5):481.

Stephen Rollnick, William R Miller, Christopher C Butler, and Mark S Aloia. 2008. Motivational interviewing in health care: helping patients change behavior. *COPD: Journal of Chronic Obstructive Pulmonary Disease* 5(3):203–203.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

Wayne F Velicer and Joseph L Fava. 1998. Affects of variable and subject sampling on factor pattern recovery. *Psychological methods* 3(2):231.

Steven R. Wilson, Rada Mihalcea, Ryan L. Boyd, and James W. Pennebaker. 2016. *Cultural influences on the measurement of personal values through words*, AI Access Foundation, volume SS-16-01 - 07, pages 314–317.

Markus Wolf, Cindy K Chung, and Hans Kordy. 2010. Inpatient treatment to online aftercare: e-mailing themes as a function of therapeutic outcomes. *Psychotherapy Research* 20(1):71–85.

Bo Xiao, Daniel Bone, Maarten Van Segbroeck, Zac E Imel, David C Atkins, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2014. Modeling therapist empathy through prosody in drug addiction counseling. In *Fifteenth Annual Conference of the International Speech Communication Association*.

Bo Xiao, Dogan Can, Panayiotis G Georgiou, David Atkins, and Shrikanth S Narayanan. 2012. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. IEEE, pages 1–4.

Bo Xiao, Zac E Imel, Panayiotis G Georgiou, David C Atkins, and Shrikanth S Narayanan. 2015. "rate my therapist": Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PloS one* 10(12):e0143055.