Modeling Therapist Empathy and Vocal Entrainment in Drug Addiction Counseling

Bo Xiao[1], *Panayiotis G. Georgiou*[1], *Zac E. Imel*[2], *David C. Atkins*[3], *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.

boxiao@usc.edu, georgiou@sipi.usc.edu, zac.imel@utah.edu
datkins@u.washington.edu, shri@sipi.usc.edu

Abstract

Empathy is considered a key curative aspect of interactive counseling based psychotherapy. In this present work, it is deemed an interpersonal behavior whereby one person communicates attention and understanding to another. The process of empathy involves "trying on" the thoughts or feelings of another person. Thus it is hypothesized to involve entrainment, wherein interlocutors become more alike in behaviors such as speech, gestures, emotions, *etc.* We extend previous algorithms on vocal similarity, measured through temporal weighting on speech features, and principal components constructed from these features. In addition, we approximate entrainment via turn-based differences in weighted pitch between speakers and turn-taking statistics. Results show these cues are significantly correlated with human ratings of empathy, and can predict therapist empathy significantly better than chance. This work establishes a link between empathy and entrainment, and proposes computational approaches to infer therapist empathy.

Index Terms: Empathy; Entrainment; PCA; Correlation; Motivational Interview

1. Introduction

Interpersonal interactions are an integral part of human life. Although common place, scientific studies of human interactions continue to pose great challenges across fields of inquiry including neuroscience [1] and psychology [2, 3]. The complex, multimodal dynamic nature of these processes, and the heterogeneity and variability in how they unfold across individuals and contexts, make it difficult for any single approach to offer complete insight into these human interaction mechanisms.

Engineering approaches offer one viable way to study human interactions and develop useful technological applications. Beyond informing designs of intuitive and natural user interfaces, computational tools and models of human interactions can inform research and practice across a variety of behavior-centered domains such as mental health. The emerging field of *Behavioral Signal Processing* [4] offers encouraging results towards such a direction: *e.g.,* developing predictive models of affective behaviors in distressed married couples' interactions using multimodal signals including acoustic [5], lexical [6], visual [7], and vocal entrainment cues [8], as well as jointly modeling both the child and the psychologist in interactive diagnostic settings for Autism [9]. The approach relies on integrating domain knowledge and engineering; *e.g.,* feature design and machine learning methods are guided by domain knowledge, and experimental results in turn validate the effects of these multimodal features and algorithms on real datasets, hence offering new insights about the interaction mechanisms. The current paper is an instantiation of our BSP work that focuses on *empathy* and *vocal entrainment* in patient-therapist interactions during counseling for drug addiction.

Empathy is described as "feeling for and taking the perspective of others". It is a basic psychological process that is evident across the phylogenetic tree (*e.g.,* rodents, apes) [10], and has been studied extensively in humans. Research on empathy has utilized a variety of measures, including non-verbal behavior, brain imaging, physiological techniques (*e.g.,* skin conductance), as well as subjective perception reported by human observers and participants [10, 11, 12]. Ratings of empathy are associated with positive outcomes in a variety of human interactions, including mother-infant dyads [13] and doctor-patient [14] interactions. In particular, empathy is considered an essential quality of therapists in psychotherapy generally and drug abuse counseling in particular. Ratings of therapist empathy are associated with treatment retention as well as positive clinical outcomes (*e.g.,* decreased substance use) [15, 16].

One of the continuing challenges of studying empathy is that it is difficult to define and quantify. However, one common theme in theoretical perspectives is its connection to similarity or entrainment in the interpersonal interaction [10, 17, 18, 19]. Entrainment [20] refers to the phenomenon where the behaviors of the interactants become more similar during the interaction. These mutually-influencing behaviors can involve, and be reflected in, a variety of cues including physiology (*e.g.,* skin conductance, heart rate, *etc.*), speaking style, movement patterns of face, body and limbs, language use and affective dynamics. Entrainment is closely related to concepts of behavior synchrony, mimicry and mirroring. Hence access to such signals offers us a path toward studying the processes underlying empathy through the vehicle of behavioral entrainment. We should note that, to study empathy in sensitive interpersonal situations such as counseling, it is critical to use easily and unobtrusively collected data to assure ecological validity; engineering advances allow for such a possibility. In sum, despite the very complex nature of empathy, the opportunity of constructing computational models through signal processing could have significant scientific and practical import.

We build upon the work of Lee *et al.* who proposed a similarity metric to assess vocal entrainment in distressed couples' conversations [8]. The measure distinguishes the potential asymmetry in entrainment between interacting participants (*e.g.,* dyads). The method constructs a Principal Component Analysis (PCA) space of the speech acoustic features (energy, spectral shape, pitch, *etc.*), and compares the energy distribution of both interlocutors' features projected onto the PCA space. In Lee's experiments [8], statistical tests suggest that the similarity is significantly higher when true couple interactions are considered compared to artificially constructed dyads. Furthermore, the similarity was found to be higher for couples associated with

This work is supported by NIH, NSF and DoD.

positive rather than negative affect.

There are other emerging works on computational modeling of behavioral similarity [21, 22] as well as therapist empathy [23]. However, to date, there have not been computational approaches aimed at modeling the relation of entrainment and empathy. Moreover, the domain of Lee *et al.*'s work on the interaction of married spouses is different from the patient-therapist interaction considered in this work. There is a fundamental role equivalence in the former, but a fundamental role inequality in the latter. In a counseling scenario, a therapist is fundamentally trying to influence the patient toward (agreed upon) behavior change. In this paper, we will tackle the problem of modeling empathy by adopting and offering novel extensions to Lee's similarity measure as well as the definition of pitch and turn taking cues. The components of the PCA space derived in Lee's approach are selectively used, and the features are temporally weighted according to their adjacency to the other interlocutor. The overview of our problem is illustrated in Fig.1.

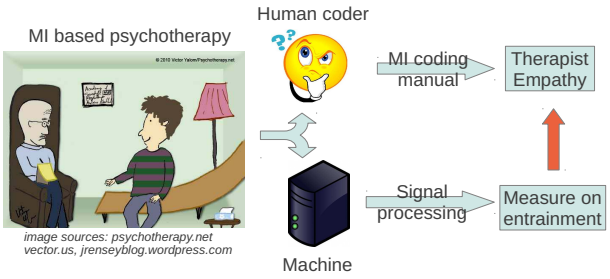

Figure 1: *Overview of the problem.*

2. Dataset

2.1. Data collection

The dataset used in our experiment comes from a counselor training study focused on a particular approach to drug addiction counseling, called Motivational Interviewing (MI) [24]. MI is a style of counseling focused on helping people to resolve ambivalence and emphasizing the intrinsic motivation of changing addictive behaviors. Empathy is hypothesized to be one of the key drivers of change in patients receiving MI [25]. In the above study 144 real therapists serving in the community participated at the beginning, with 123 completing the whole process. Therapists had a mean age of 46.1±11.6 years, a mean clinical service of 9.5±8.4 years, where 70% were female. Three *Standardized Patient* (SP) actors role played clients in about half of all the recordings, while the rest were real clients. Each participating therapist interviewed one or two SPs. Each session is about 20 min long recorded with single channel far field microphone. At collection time the intended consumers were human annotators and as such audio quality is challenging for machine processing.

Three human coders evaluated the recordings using a specially designed coding system, the *Motivational Interviewing Treatment Integrity* (MITI) [25]. As a result each session received a global rating of empathy on a Likert scale (discrete) from 1 to 7. Each of the coders received 40 hours training and joined weekly one hour discussion during the 18 months of study. Inter-rater reliability assessed via *Intra-Class Correlation* (ICC) has a mean of 0.67±0.16, while ICC for the same coder over time has a mean of 0.79±0.13. For all 182 sessions that were coded twice, the correlation of empathy scores is 0.87. No session was triple-coded.

2.2. Data pre-processing

For the present study, we consider sessions on the two extremes of empathy rating (mean value if double coded) involving a SP.

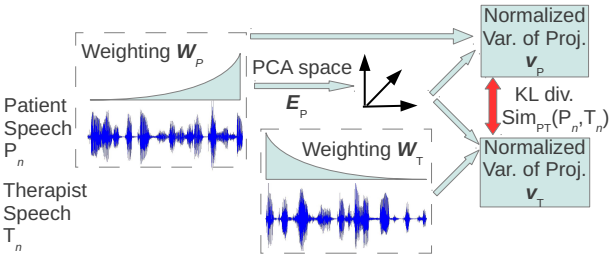

Figure 2: *Procedure for computing similarity measure.*

In total 119 sessions were included, where 72 sessions were on the high-empathy end with ratings of range 5-7 and mean 6.04±0.65, and 47 sessions were on the low-empathy end with range 1-3 and mean rating 2.17±0.55. We excluded sessions with an empathy rating of 4. Automatic voice activity detection using the "SHoUT" toolkit [26] was applied to all sessions, followed by automatic diarization and manual correction into therapist's speech and client's speech. Since we want to analyze the similarity between two interlocutors, overlapped speech was excluded from processing.

We extracted 14 dimensional Mel Frequency Cepstral Coefficients (MFCCs) including 0-*th* dimension, with 25 ms window, 10 ms shift and cepstral liftering. We extracted pitch using the method of subharmonic to harmonic ratio [27]. In order to filter out pitch doubling and halving, we first eliminated pitch values below 55 Hz, then located the central pitch by finding the mode of the result. We created a three component Gaussian mixture model initialized at the central pitch, twice its value and 1/2 of it. Any sample having a maximum posterior probability from the later two components were divided or multiplied by 2 accordingly. Finally, both of the MFCCs and pitch features were Z-normalized per session per speaker.

3. Extraction of Speech Acoustic Cues

3.1. Acoustic similarity measure

In this section we introduce the framework of our vocal similarity measure. We compare speech segments on a per-turn basis, *i.e.,* every turn of a speaker is compared to the nearest previous turn from the other speaker (so as to maintain causality). If the total duration of speech in either the current speaker or the previous speaker's turn is less than 2 sec, we skip the current speaker's turn since the amount of speech frames might be sparse for statistical stability. Note that we consider turn boundaries only at speaker changes but not pauses within a speaker, and silence frames are not counted hereafter.

Without losing generality, let us consider speaker B's turns following speaker A's turns. Let A have N turns in consideration, denoted $\{A_n\}_{n=1}^{N}$, where the corresponding B turns are $\{B_n\}_{n=1}^{N}$. Let us define the similarity measure between A_n and B_n as $\texttt{sim}_{AB}(A_n, B_n)$, which is computed as follows, and illustrated in Fig.2.

3.1.1. Construction of PCA space

Let $\mathbf{F}_A$ be the matrix of zero-mean acoustic features in A_n, in L_A^n rows (frames) and K columns (features). We associate a weight $w_A^i = \theta^{L_A^n - i}$ with the i-*th* feature vector, where $i = 1, 2, \cdots, L_A^n$, and $0 < \theta < 1$ is a forgetting factor. Such time reversed weighting assumes that the speech closer to the end of a turn may affect the other interlocutor more. Let $\mathbf{W}_A = \texttt{diag}(w_A^1, w_A^2, \cdots, w_A^{L_A^n})$ be the weighting matrix. We obtain the weighted covariance matrix [28] as (1):

$$\mathbf{C} = \frac{1}{Z_A}\mathbf{F}_A^{'}\mathbf{W}_A^{'}\mathbf{W}_A\mathbf{F}_A, \quad \text{where } Z_A = \sum_{i=1}^{L_A^n}(w_A^i)^2 \qquad (1)$$

Let $\mathbf{E}_A$ be the matrix of eigenvectors of $\mathbf{C}$, where each column is an eigenvector sorted by the descending order of eigenvalues. This gives the description of interlocutor A's PCA space.

3.1.2. Projection to PCA space

For the purpose of comparison, we project both A and B's acoustic features to the PCA space $\mathbf{E}_A$. Let $\mathbf{F}_B$ be the zero-mean feature matrix of B. We also associate a weight $w_B^i = \theta^{i-1}$ to each row of $\mathbf{F}_B$, where $i = 1, 2, \cdots, L_B^n$. Note here the weighting assumes that the speech closer to the beginning of a turn may be more affected by the other interlocutor. Let $\mathbf{W}_B$ be the weighting matrix for B. We obtain the projection of A as $\mathbf{X}_A = \mathbf{W}_A\mathbf{F}_A\mathbf{E}_A$, and the projection of B as $\mathbf{X}_B = \mathbf{W}_B\mathbf{F}_B\mathbf{E}_A$. The distribution of energy towards every PCA component is reflected in the weighted variance of each column in $\mathbf{X}_A$ and $\mathbf{X}_B$, where we define the weighted variance as (2):

$$\mathbf{v}_A(k) = \frac{1}{Z_A}\sum_{i=1}^{L_A^n}(w_A^i\mathbf{X}_A(i,k))^2, k = 1, 2, \cdots, K \quad (2)$$

$\mathbf{v}_B$ is obtained in the same way. We further normalize $\mathbf{v}_A$ and $\mathbf{v}_B$ by their respective vector sum, so that they can be treated numerically as a probability distribution.

3.1.3. Similarity measure as KL divergence

We can compare $\mathbf{v}_A$ and $\mathbf{v}_B$ to obtain a similarity measure between the interlocutor utterances. However, it is not necessary to use the full signal dimensionality. The main components in PCA might be more relevant and may capture the most important signal trends in the observed window. Assume we keep the first J dimensions, $2 \le J \le K$, so that $\widetilde{\mathbf{v}_A} = \frac{\mathbf{v}_A(1\cdots J)}{\sum \mathbf{v}_A(1\cdots J)}$, and so as $\widetilde{\mathbf{v}_B}$. The symmetric Kullback-Leibler divergence is defined as in (3).

$$D(p||q) = \frac{1}{2}\sum_{j=1}^{J} p(j)\texttt{log}\frac{p(j)}{q(j)} + \frac{1}{2}\sum_{j=1}^{J} q(j)\texttt{log}\frac{q(j)}{p(j)} \quad (3)$$

Finally, the similarity measure $\texttt{sim}_{AB}(A_n, B_n)$ is defined as in (4). For a session-wise similarity metric, we use the mean and variance of $\texttt{sim}_{AB}(A_n, B_n)$ for all n, denoted $M_{\text{sim}}(AB, AB)$ and $V_{\text{sim}}(AB, AB)$.

$$\texttt{sim}_{AB}(A_n, B_n) = D(\widetilde{\mathbf{v}_A}||\widetilde{\mathbf{v}_B}) \quad (4)$$

3.1.4. Other variations

There are three variations to $\texttt{sim}_{AB}(A_n, B_n)$. First, one can construct the PCA space $\mathbf{E}_B$ from the weighted features of B, then project both $\mathbf{F}_A$ and $\mathbf{F}_B$ onto that space, and obtain $\texttt{sim}_{AB}(B_n, A_n)$. Additionally, one can swap the order of A and B, *i.e.,* consider A's turns following B's turns, then repeat the previous two ways. This yields $\texttt{sim}_{BA}(B_n, A_n)$ and $\texttt{sim}_{BA}(A_n, B_n)$. The session level similarity metrics are denoted in a similar fashion, *e.g.,* $M_{\text{sim}}(BA, AB)$ represents the mean of the $\texttt{sim}_{BA}(A_n, B_n)$, *etc.*

3.2. Average pitch cues

In addition to the measures above, some simpler cues that offer a more direct physical meaning may also be useful. One cue is the absolute difference of mean weighted pitch. Frames without a valid (non-zero) pitch value are not counted, yet the weighting is still applied on the whole time axis. Therefore, we denote pitch as p, and the indicator function as $I(p)$, which is 1 if the condition $p \neq 0$ is true, and 0 otherwise. The difference of pitch is symmetric for a pair of turns. For example, let $\delta p_{AB}(A_n, B_n)$ denote the absolute difference of mean weighted pitch when B follows A's turn. Let $\mathbf{p}_A$, $\mathbf{p}_B$ be the vector of pitch in the n-*th* turn for A and B, respectively. $\delta p_{AB}(A_n, B_n)$ is defined as in (5).

$$\begin{aligned}\delta p_{AB}(A_n, B_n) &= |\overline{p_A} - \overline{p_B}| \quad \text{where} \quad (5)\\ \overline{p} &= \frac{\sum_{i=1}^{L} I(p(i))w^i p(i)}{\sum_{i=1}^{L} I(p(i))w^i},\end{aligned}$$

One can obtain $\delta p_{BA}(B_n, A_n)$ in the same way, with A following B's turns and the weighting exchanged accordingly. For the session level metric, we again take the mean (*e.g.,* $M_{\text{pit}}(BA)$ for the mean of $\delta p_{BA}(B_n, A_n)$) and variance (*e.g.,* $V_{\text{pit}}(BA)$) of the above turn level measures.

3.3. Turn taking cues

Moreover, directly from the turn taking information, we collect two session level ratios, which may reflect the empathy of the therapist specifically in the motivational interviewing scenario. The first one is the ratio of patient speech time over the total speech time, denoted R_t. The second one is the ratio of the count of patient speaking segments (separated by pauses and speaker changes) over the total number of speaking segments in the session, denoted R_s.

3.4. Integrated measure

We have proposed above three classes of cues that we believe relate to entrainment and likely to empathy. The similarity and pitch measures are also dependent on a range of parameters and choices: specifically speaker order (*e.g.,* $M_{\text{sim}}(AB, AB)$ *vs.* $M_{\text{sim}}(BA, BA)$) and the numerical choices of the PCA dimension (J) and forgetting factor (θ). These cues may be highly correlated in value, and also complementary in reflecting empathy. Therefore, we want to design an integrated measure y fusing the above cues to achieve stronger correlation.

Given a training set of m sessions with empathy score $\mathbf{e}$, we first compute the similarity, pitch and turn taking cues, collected as $\mathbf{U}$ with m rows and c columns, where c is the total count of cues. We normalize by removing the per-cue mean (M_U) to get $\overline{\mathbf{U}}$. Second, using $\overline{\mathbf{U}}$ we derive its PCA space $\mathbf{E}_U$ and projection $\mathbf{X}_U = \overline{\mathbf{U}}\mathbf{E}_U$. Third, we select a single component (column) $\mathbf{E}_U(p)$ in $\mathbf{E}_U$ by maximizing the correlation with empathy, *i.e.,* $p = \arg\max_{p'} \texttt{corr}(\mathbf{X}_U(p'), \mathbf{e})$, where $1 \le p' \le c$. Finally, for a test session with cue vector $\mathbf{u}$, we obtain $y = (\mathbf{u} - M_U)\mathbf{E}_U(p)$.

4. Experiments

4.1. Correlation by an individual cue

We first examine the various cues by correlating them with the empathy ratings, and check the p-value with student's t test. Let T stand for therapist and P for patient. For the similarity measure we use a 15-dimensional feature composed of pitch and 14 MFCCs (including 0-*th* order coefficient). Under a significance level of 0.01, we observe that $M_{\text{sim}}(PT, PT)$ (T following P's turn, constructing PCA from P) with $J = 2$ yields high negative correlation. Similarly the mean pitch $M_{\text{pit}}(PT)$ and pitch variance $V_{\text{pit}}(PT)$ measures are highly correlated with empathy. The forgetting factor θ has an effect but results are not very sensitive to θ changes. The effect of J will be discussed in Sec. 5. The negative sign of correlation is expected, which suggests that smaller divergence, *i.e.,* stronger similarity, is associated with higher empathy. R_t and R_s both have positive significant correlations, suggesting more patient talk is associated with higher therapist empathy. In Table 1 we show the correlation in the above cases, where θ is chosen such that the attenuation at 30 sec is 1 (no weighting), 10^{-1}, 10^{-2} or 10^{-3}. The other variations of M_{sim}, V_{sim}, M_{pit}, V_{pit} are not significant.

θ^{3000}	1	10^{-1}	10^{-2}	10^{-3}
$M_{sim}(PT,PT)$	-0.24	-0.25	-0.27	-0.29
$V_{sim}(PT,PT)^*$	-0.22	-0.23	-0.21	-0.19
$M_{pit}(PT)$	-0.31	-0.32	-0.31	-0.30
$V_{pit}(PT)$	-0.32	-0.34	-0.35	-0.34
R_t	0.27			
R_s	0.28			

* significant at p-value 0.05

Table 1: *Correlation of various cues and therapist empathy.*

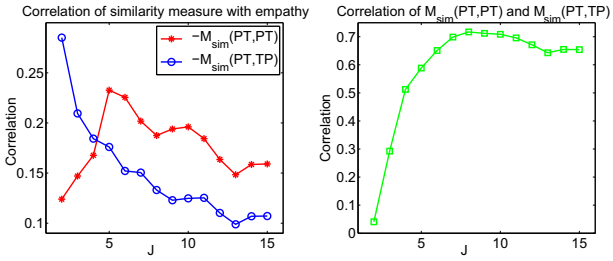

Figure 3: *The effect of J on correlation with empathy.*

4.2. Correlation by integrated measure

Although we have seen the correlations, in a rigorous sense of training and testing, from all variations of cues, we select those having significant correlations (p-value < 0.05) based on the training set. Due to the very limited number of sessions, we conduct this process through a leave-one-session-out cross-validation. In each fold we compute the integrated measure $\mathbf{X}_U(p)$ as in Sec. 3.4, for the training sessions, and compute the integrated measure y for the test session. Using the trained models for $\mathbf{X}_U(p)$ and $\mathbf{e}$, we construct a linear regression model to estimate the empathy rating of the test session. In parallel we apply a logistic regression to predict the "high" *vs.* "low" empathy (see Sec. 2.2) trained with the binarized empathy ratings. In Table 2 we report the averaged absolute correlation (may be positive or negative) of $\mathbf{X}_U(p)$ and $\mathbf{e}$ in training, and correlation of estimated empathy and empathy ratings in testing, as well as the binary prediction accuracy in both training and testing. The testing accuracy is significantly ($p < 0.05$) better than chance level (0.605 based on the prior).

5. Discussion

5.1. Asymmetry induced by J

In Sec. 4 we employed $J = 2$ having observed it provides the best performance. We further observed an asymmetry between $M_{\text{sim}}(PT,PT)$ and $M_{\text{sim}}(PT,TP)$. In order to analyze these observations, we compute these two measures with J from 2 to 15. The correlation of the two measures is shown in Fig.3. The plot demonstrates that with small J, similarities induced by the *renormalized* high ranked PCA components from $\mathbf{E}_P$ and $\mathbf{E}_T$ are very different; as J moves higher, the comparison is more comprehensive so that the two measures behave more alike. However, the correlation with empathy drops when more components are considered. We want to investigate and explain in future work why a small ($J = 2$) PCA dimensionality leads to higher correlation.

Training		Testing	
Corr.	Acc.	Corr.	Acc.
0.47 ± 0.01	0.71 ± 0.01	0.43	0.70*

$*$p-value = 0.02 in binomial test

Table 2: *Performance in cross-validation.*

θ^{3000}	1	10^{-3}
$M_{\text{sim}}(PT,PT)$ & $M_{\text{sim}}(TP,PT)$	0.85	0.66
$M_{\text{pit}}(PT)$ & $M_{\text{pit}}(TP)$	0.71	0.53

Table 3: *Correlation of cues in different turn order ($J = 2$).*

5.2. Asymmetry of turn order

The performance also differs with respect to turn order. We compute the correlation of cues derived in different turn order in Table. 3. This demonstrates the difference between these measures – that they indeed measure a different phenomenon – despite their similarity, and it suggests that the asymmetry of performance is induced by the asymmetry of the participant's role, and the coder's attention focused on the therapist. We intend to investigate the directionality of entrainment and that of the measures in future work.

5.3. Comparison with behavior counts

Lacking a comparable approach in literature, we look at the manually collected behavior counts in the MITI coded corpus, which are hypothesized to reflect therapist empathy and are specified by domain experts. Out of the 7 behavior counts, significant ones (correlation with empathy) are "complex reflection" (0.70), "MI non-adherent" (-0.61), "closed question" (-0.48), "simple reflection" (0.35) and "giving information" (-0.33). In MITI coding manual [25], coders are expected to get a *gestalt* (a unified whole) impression of therapist empathy. Although how exactly the coder fuses perspective information into a numerical value (a cognitive process) is unclear, the correlations imply that one can still infer therapist empathy through his/her own behavior. The correlation achieved by computational approach is in the range of that obtained by more abstract cues that are identified by trained human coders.

5.4. Other measures

We also attempted to use the "global" and "symmetric" version of similarity measure in [8], but the correlation was weak. It might suggest difference of application – couples' conversation where the interlocutors had been interacting much longer *vs.* therapists and patients in MI sessions meeting for the first time – has a fundamental effect on entrainment behavior.

6. Conclusion

In this paper we investigated the statistical relationship between empathy and vocal entrainment, two important human interaction characteristics. Perceived empathy was rated by domain experts while vocal entrainment was estimated by Behavioral Signal Processing algorithms. Particularly in the psychotherapy scenario of motivational interviewing, the link between entrainment and empathy was established both theoretically and empirically, and now preliminarily with the proposed computational approach. By extending previous measures on vocal entrainment, and incorporating new pitch similarity cues as well as turn taking statistics, we showed that not only are these cues significantly correlated with empathy individually, but also their integrated metric has even higher correlation with empathy, comparable with that obtained by manually collected behavioral information. This result verifies the link between entrainment and empathy, and also proposes computational ways to infer them.

From our analysis we observed asymmetric performance among the measure variations and we intend to study that further in future work. This work raises a lot of interesting questions for both the engineering and psychology experts. For instance, on the engineering side BSP will aim to improve the current entrainment metrics and analyze their variants while from the psychology perspective we would like to better understand entrainment asymmetry and directionality.

7. References

[1] M. Iacoboni, *Mirroring people: The science of empathy and how we connect with others*. Picador, 2009.

[2] M. Knapp and J. Hall, *Nonverbal Communication in Human Interaction*, 7th ed. Boston: Wadsworth, Cengage Learning, 2007.

[3] J. Harrigan, R. Rosenthal, and K. Scherer, *The new handbook of Methods in Nonverbal Behavior Research*. New York: Oxford University Press, 2005, pp. 137–198.

[4] S. Narayanan and P. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceeding of IEEE*, 2012.

[5] M. Black, A. Katsamanis, B. Baucom, C. Lee, A. Lammert, A. Christensen, P. Georgiou, and S. Narayanan, "Toward automating a human behavioral coding system for married couples interactions using speech acoustic features," *Speech Communication*, 2011.

[6] P. Georgiou, M. Black, A. Lammert, B. Baucom, and S. Narayanan, ""that's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?" in *Proc. ACII*, 2011, pp. 87–96.

[7] B. Xiao, P. Georgiou, B. Baucom, and S. Narayanan, "Data driven modeling of head motion towards analysis of behaviors in couple interactions," in *Proc. ICASSP*, May 2013.

[8] C. Lee, A. Katsamanis, M. Black, B. Baucom, A. Christensen, P. Georgiou, and S. Narayanan, "Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions," *Computer Speech & Language*, 2012.

[9] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Proc. Interspeech*, 2012.

[10] S. D. Preston and F. De Waal, "Empathy: Its ultimate and proximate bases," *Behavioral and Brain Sciences*, vol. 25, no. 01, pp. 1–20, 2002.

[11] M. Iacoboni, "Imitation, empathy, and mirror neurons," *Annual review of psychology*, vol. 60, pp. 653–670, 2009.

[12] C. D. Marci, J. Ham, E. Moran, and S. P. Orr, "Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy," *The Journal of nervous and mental disease*, vol. 195, no. 2, pp. 103–111, 2007.

[13] N. D. Feshbach, "Parental empathy and child adjustment / maladjustment," *Empathy and its development*, p. 271, 1990.

[14] P. Bellet and M. Maloney, "The importance of empathy as an interviewing skill in medicine," *Journal of the American Medical Association*, vol. 266, no. 13, pp. 1831–1832, 1991.

[15] R. Elliott, A. C. Bohart, J. C. Watson, and L. S. Greenberg, "Empathy," *Psychotherapy*, vol. 48, no. 1, p. 43, 2011.

[16] W. R. Miller and G. S. Rose, "Toward a theory of motivational interviewing," *American Psychologist*, vol. 64, no. 6, p. 527, 2009.

[17] J. Decety and P. Jackson, "The functional architecture of human empathy," *Behavioral and cognitive neuroscience reviews*, vol. 3, no. 2, pp. 71–100, 2004.

[18] T. Arizmendi, "Linking mechanisms: Emotional contagion, empathy, and imagery," *Psychoanalytic Psychology*, vol. 28, no. 3, p. 405, 2011.

[19] J. B. Bavelas, A. Black, C. R. Lemery, and J. Mullett, "Motor mimicry as primitive empathy," *Empathy and its Development*, p. 317, 1990.

[20] T. Wheatley, O. Kang, C. Parkinson, and C. Looser, "From mind perception to mental connection: Synchrony as a mechanism for social understanding," *Social and Personality Psychology Compass*, vol. 6, no. 8, pp. 589–606, 2012.

[21] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Transactions on Affective Computing*, 2012.

[22] B. Xiao, P. Georgiou, C. Lee, B. Baucom, and S. Narayanan, "Head motion synchrony and its correlation to affectivity in dyadic interactions," in *Proc. ICME*, Jul. 2013.

[23] B. Xiao, D. Can, P. G. Georgiou, D. Atkins, and S. S. Narayanan, "Analyzing the language of therapist empathy inmotivational interview based psychotherapy," in *APSIPA ASC*, Dec. 2012.

[24] J. S. Baer, E. A. Wells, D. B. Rosengren, B. Hartzler, B. Beadnell, and C. Dunn, "Agency context and tailored training in technology transfer: A pilot evaluation of motivational interviewing training for community counselors," *Journal of substance abuse treatment*, vol. 37, no. 2, p. 191, 2009.

[25] T. Moyers, T. Martin, J. Manuel, and W. Miller, "The motivational interviewing treatment integrity (miti) code: Version 2.0," *Unpublished. Albuquerque, NM: University of New Mexico, Center on Alcoholism, Substance Abuse and Addictions*, 2008.

[26] M. A. H. Huijbregts, "Segmentation, diarization and speech transcription: surprise data unraveled," Ph.D. dissertation, University of Twente, 2008.

[27] X. Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," in *Proc. ICASSP*, vol. 1. IEEE, 2002, pp. 333–336.

[28] H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "A general framework for increasing the robustness of pca-based correlation clustering algorithms," in *Scientific and Statistical Database Management*. Springer, 2008, pp. 418–435.