

## BRIEF REPORT

# The Association of Therapist Empathy and Synchrony in Vocally Encoded Arousal

Zac E. Imel, Jacqueline S. Barco, Halley J. Brown,  
and Brian R. Baucom  
University of Utah

John S. Baer  
University of Washington and VA Puget Sound Healthcare  
System–Seattle Division

John C. Kircher  
University of Utah

David C. Atkins  
University of Washington

Empathy is a critical ingredient in motivational interviewing (MI) and in psychotherapy generally. It is typically defined as the ability to experience and understand the feelings of another. Basic science indicates that empathy is related to the development of synchrony in dyads. However, in clinical research, empathy has proved difficult to operationalize and measure, and has mostly relied on the *felt sense* of observers, clients, or therapists. We extracted estimates of therapist and standardized patient (SP) vocally encoded arousal (mean fundamental frequency; mean  $f_0$ ) in 89 MI sessions with high and low empathy ratings from independent observers. We hypothesized (a) therapist and SP mean  $f_0$  would be correlated and (b) the correlation of therapist and SP mean  $f_0$  would be greater in sessions with high empathy as compared with low. On the basis of a multivariate mixed model, the correlation between therapist and SP mean  $f_0$  was large ( $r = .71$ ) and close to 0 in randomly assigned therapist–SP dyads ( $r = -.08$ ). The association was higher in sessions with high empathy ratings ( $r = .80$ ) than in sessions with low ratings ( $r = .36$ ). There was strong evidence for vocal synchrony in clinical dyads as well as for the association of synchrony with empathy ratings, illustrating the relevance of basic psychological processes to clinical interactions. These findings provide initial evidence for an objective and nonobtrusive method for assessing therapist performance. Novel indicators of therapist empathy may have implications for the study of MI process as well as the training of therapists generally.

*Keywords:* empathy, synchrony, speech signal processing

*Empathy* is an interpersonal process typically defined as the ability to both understand and experience the feelings of another person (Preston & De Waal, 2002). Within the field of psycho-

therapy, empathy is thought to be important across treatments but has been specifically emphasized within motivational interviewing (MI). *MI* refers to a class of evidence-based psychotherapies that specifies a particular linguistic approach to treatment wherein the therapist promotes client “change talk” while maintaining a non-judgmental and empathic stance. In MI, empathy is hypothesized to have both direct effects on treatment outcomes and indirect effects via the facilitation of client change talk (Miller & Rose, 2009). These predictions are consistent with evidence that ratings of therapist empathy are related to client outcomes in MI (Miller & Rose, 2009; Moyers & Miller, 2013) and across psychotherapies (Elliott, Bohart, Watson, & Greenberg, 2011).

However, the empirical study of empathy has proved challenging, beginning with its definition. In clinical research, empathy is treated as a *felt sense* construct that relies on human perceptual skill for evaluation. Although there is no single definition, there is some consensus that empathy is composed of three processes: (a) emotional simulation—mirroring of the others experience, (b) perspective taking—understanding the client, and (c) emotion regulation—soothing interpersonal distress (Eisenberg & Eggum, 2009). Most clinical research has emphasized perspective-taking processes (Elliott et al., 2011), and corresponding definitions can be

---

This article was published Online First November 25, 2013.

Zac E. Imel, Jacqueline S. Barco, and Halley J. Brown, Department of Educational Psychology, University of Utah; Brian R. Baucom, Department of Psychology, University of Utah; John S. Baer, Department of Psychology, University of Washington, and VA Puget Sound Healthcare System–Seattle Division; John C. Kircher, Department of Educational Psychology, University of Utah; David C. Atkins, Department of Psychiatry and Behavioral Sciences, University of Washington.

Funding for the preparation of this article was provided by a University of Utah Seed Grant and by National Institute on Drug Abuse (NIDA) of the National Institutes of Health Award R34/DA034860 as well as by National Institute on Alcohol Abuse and Alcoholism Award R01/AA018673. The original trial was supported by NIDA Grant R01/DA016360. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Correspondence concerning this article should be addressed to Zac E. Imel, Counseling Psychology Program, Department of Educational Psychology, University of Utah, 1705 Campus Center Drive, Room 327, Salt Lake City, UT 84112. E-mail: zac.imel@utah.edu

abstract and difficult to operationalize—“entering the private perceptual world of the other” (Rogers, 1980, p. 142). The Motivational Interviewing Treatment Integrity Scale (MITI), a standard MI behavioral coding manual, defines empathy as whether the therapist is able “to ‘try on’ what the client feels or thinks.” A low rating (1) indicates that the therapist showed no interest in the client’s perspective, and a high rating (7) indicates that the therapist demonstrated “a deep understanding of client’s point of view” (Moyers, Martin, & Manuel, 2005, p. 14). Moreover, in MI, empathy is most often quantified through behavioral coding systems, which can be extremely time-consuming (e.g., Moyers, Martin, Manuel, Hendrickson, & Miller, 2005), and reliability estimates can often be quite low (ICCs of approximately .40; Moyers, Miller, & Hendrickson, 2005; Vader, Walters, Prabhu, Houck, & Field, 2010; see also Moyers, Martin, Catley, Harris, & Ahluwalia, 2003, for an exception, ICC = .77). Low and variable estimates of reliability are consistent with the idea that empathy is not well defined operationally.

It may be that there is a shared, intuitive understanding of what empathy is, but a lack of behavioral specificity and practical problems inherent in the behavioral coding make it challenging to evaluate. Methods for evaluating therapist empathy are needed to provide more specific information about what happens in therapy dyads.

### Empathy and Synchrony

In addition to perspective-taking definitions, empathy has also been conceptualized as a general process that involves some form of mirroring (Preston & De Waal, 2002). For example, the perception action model defines empathy as a process wherein a subject’s state results from the attended perception of the object’s state (Hoffman, 2000; Preston & De Waal, 2002). Here, empathy depends on a process of imitation or synchrony wherein humans learn how others feel by experiencing a representation of a similar state (Iacoboni, 2009), analogous to the MI therapist “trying on” what the client feels.

There is broad evidence for synchrony of behavior in human dyads across a variety of outcomes (e.g., words, acoustic features of speech; see also communication accommodation theory; Chartrand & van Baaren, 2009; Giles & Ogay, 2007; Gregory & Hoyt, 1982; Lee et al., 2011; Pickering & Garrod, 2004). Quantitative linguistic research has established word use matching in the laboratory as well as naturalistic interactions (e.g., Watergate tapes; Niederhoffer & Pennebaker, 2002; see also Ireland & Pennebaker, 2010). With respect to vocal acoustics, the most commonly studied feature is *mean fundamental frequency* (mean  $f_0$ ), which refers to the vibration created by the vocal folds in the throat and corresponds to the lowest harmonic produced during speech (Kappas, Hess, & Scherer, 1991). Mean  $f_0$  is highly correlated with perceived pitch (up to  $r = .9$ ) and is widely interpreted as a measure of vocally encoded emotional arousal (Juslin & Scherer, 2005; Kappas et al., 1991; Russell, Bachorowski, & Fernández-Dols, 2003). *Vocally encoded emotional arousal* refers to the degree of emotional activation conveyed by the voice (e.g., high = excited, angry, or nervous; low = bored, calm, or content), and higher levels of  $f_0$  have been linked to higher levels of physiological (e.g., higher heart rate, higher systolic and diastolic blood pressure, and greater cortisol output; Scherer, 1989; Weusthoff, Baucom, &

Hahlweg, 2013) and self-reported emotional arousal (e.g., Baucom et al., 2012). A series of studies indicate that  $f_0$  converges as dyads converse (see Gregory, 1990; Gregory, Webster, & Huang, 1990).

Beyond the general phenomenon of dyadic interactions, evidence points to a relationship between dyadic synchrony and the quality of interpersonal interactions both in clinical (therapist–client) and nonclinical dyads (Dimascio, Boyd, & Greenblatt, 1957; Lee et al., 2012; Levenson & Ruef, 1992; Robinson, 1982). For example, the more participants in a social psychology experiment imitated the nonverbal behavior of confederates, the higher they scored on a self-reported measure of empathy (Chartrand & Bargh, 1999), and word use matching was associated with relationship initiation and stability in speed dates (Ireland et al., 2011). In a recent study of supportive behavior in couples, there was evidence that the similarity of partner self-report measures of emotional arousal predicted more skillful provision of support (Verhofstadt, Buysee, Ickes, Davis, & Devoldre, 2008). Clinically, patient and therapist body movement synchrony was higher in genuine clinical interactions as compared with controls and was associated with higher ratings of relationship quality and treatment outcome (Ramseyer & Tschacher, 2011). In psychodynamic psychotherapy, the coordination of client and therapist skin conductance, a physiological measure of arousal, was correlated with client-perceived therapist empathy (Marci, Ham, Moran, & Orr, 2007).

The above studies provide reason to suspect that behavioral and physiological synchrony is one mechanism for the communication of therapist empathy. However, many physiological measures can be obtrusive and may not be practical in many clinical or research settings. In addition, MI and psychotherapy generally are verbally mediated treatments. Although physiological and nonverbal measures provide important information, the quality of therapist empathy should also be found in the conversation that occurs between therapists and clients. Specifically, it is not sufficient for a therapist to simply experience an internal representation of the client’s experience. The empathic therapist should also communicate his or her response through reflecting the client’s communication in a way that highlights the client’s experience.

At present, there are no studies of vocal synchrony in psychotherapy or whether it reflects therapist empathy. However, measures of vocal acoustic like mean  $f_0$  hold promise for evaluating psychotherapy (Rochman & Amir, 2013). For example, mean levels of  $f_0$  range during a precouples therapy interaction predicted treatment outcomes 2 years posttherapy (Baucom, Atkins, Simpson, & Christensen, 2009) and various measures of vocal acoustics discriminated between two analogue forms of psychotherapy (higher  $f_0$  during an empty chair exercise; Diamond, Rochman, & Amir, 2010).

### Summary and Current Study

The importance of therapist empathy is clear, but current metrics are labor intensive and rely on abstract definitions without clear ties to theories of the empathic process. There is evidence to suggest that synchrony may be an indicator of empathy, but empirical research has yet to be conducted on basic acoustic processes related to empathy during psychotherapy. We used speech signal processing methods to examine synchrony in mean  $f_0$  and its association with coded empathy in 89 MI sessions

conducted with standardized patients (SPs). We had two primary hypotheses. First, we predicted that there would be therapist–SP synchrony in vocally encoded arousal within MI sessions. Second, we predicted that greater synchrony would be associated with higher independent observer ratings of therapist empathy.

## Method

### Data Source

Data were obtained from an eight-site<sup>1</sup> training study in which providers were recruited from National Institute of Drug Abuse-affiliated community substance abuse treatment facilities in the state of Washington (Baer et al., 2009; see also Imel et al., 2013). Therapists were offered free MI training, continuing education credits, and financial compensation for their participation. The original study randomized participating providers to one of two MI training formats: either (a) a 2-day workshop or (b) an ongoing training that included five 2-hr training sessions between which learners practiced with SPs, immediately received feedback from their SPs, and then had their audiotapes further reviewed by trainers who provided those learners with written feedback. Each group received approximately 15 total hours of time with MI trainers. Provider's MI skills were evaluated on the basis of six MI treatment sessions conducted with both SPs and real patients (RPs) at pretraining, posttraining, and 3-month follow-up. Each session was 20 min in total length. A provider submitted one SP and one RP session recording at each time point. Results showed a significant effect of training, with provider MI skills increasing from pre- to posttraining and maintained at follow-up, but no differences due to training format. There was a large correlation between empathy ratings in SP and RP sessions ( $r = .75$ ; Imel et al., 2013; see also Baer et al., 2009, for additional details on study design).

**Selected sessions.** The data included digital audio and observer ratings of empathy from 89 selected MI sessions (drawn from a total of 485 SP sessions—18% of sessions—one session per therapist). Manual audio segmentation is labor intensive, requiring 1–2 hr per session (i.e., identifying speakers within a session; see the Vocal Fundamental Frequency section, below), and we segmented sessions rated in the top or bottom quartile (empathy rating of 1 or 2 “low” or a 6 or 7 “high”). If a therapist had multiple sessions, we randomly selected one session. Session recordings also needed to be of sufficient quality to segment and extract mean  $f_0$  values. Recordings obtained from busy substance abuse treatment clinics can introduce ambient noise, occasional background chatter, and reverberation effects. Accordingly, 31 sessions were excluded because recordings were not sufficient. Two sessions were excluded because of a corrupted source file. After the above data reduction, 89 SP sessions were used for the current study, including 54 high-empathy sessions and 35 low-empathy sessions.<sup>2</sup>

The selected sessions included 89 therapists (selected from a total of 189 therapists, 47% of therapists). Therapists were 85% female, and 83% Caucasian, 9% African American, 6% Native American, 6% Hispanic/Latino, 4% Asian, and 3% Pacific Islander. Mean age was 45.9 years ( $SD = 11.0$ ), and mean length of experience in clinical service provision was 10.7 years ( $SD = 9.7$ ).

Twenty-nine percent had completed a graduate degree, 24% a bachelor's degree, 32% an associate's degree, and 10% a high school diploma or equivalent. A majority of the sample (57%) reported some prior exposure to MI through training workshops, coursework, reading of MI texts or journal articles, or other means before study trainings.

**SPs.** One of three female SPs saw therapists at each time point. SPs portrayed a recently referred client with a substance use problem and characteristics common for agency clientele. There was no standardized protocol for assignment of SPs across all sites. SPs were trained to provide realistic clinical vignettes and to respond flexibly to counselors during sessions. SP training was directed by study investigators and involved considerable role-play practice in which SPs were provided feedback on authenticity of their performances. Therapists were aware that SPs were confederates. Research using SPs can provide more reliable estimates of provider behavior than RPs (e.g., with RPs therapist differences may reflect patient caseload heterogeneity) and also avoid problems with missing data and audio quality that are common when samples of clinical work are requested from therapists in the community (Baer et al., 2004). Empathy ratings did not differ across the three SPs,  $F(2, 87) = 0.082$   $p > .5$ .

### Measures

**MITI 3.0 (Moyers, Martin, & Manuel, 2005).** We used the global empathy rating from the MITI, a standard measure of therapist fidelity in MI. Ratings of therapist empathy were obtained from audio recordings of sessions via a single rating of the session, using a 7-point Likert scale (see the introduction for prompt and scale anchors). Raters were trained and supervised in the MITI and were blind to assessment timing and practitioner identifiers. The interreliability of empathy codes was in the upper range of previous studies ( $ICC = 0.60$ ). As therapists were not trained to a certain level of competence before participation, an advantage of the current data set is that scores spanned the full range of the scale (1–7). Typical MI trials have skewed empathy scores with very few low ratings. In the current set of 89 sessions, over six standard deviations separated the means for sessions rated as high ( $M = 6.1$ ,  $SD = 0.67$ ) or low empathy ( $M = 2.2$ ,  $SD = 0.59$ ).

**Vocal fundamental frequency.** Mean  $f_0$  was used to assess therapist and SP vocally encoded arousal. Often described as vocal

<sup>1</sup> The majority of the data were collected at specific sites. There were six primary study sites and two pilot sites (for eight total). In addition, there was an open enrollment phase of the project wherein therapists were not nested within sites. In theory, it might be possible to examine the subset of data wherein therapists were nested within sites to determine whether therapists and SPs at particular sites reliably varied in their mean  $f_0$ . However, sites also varied notably in ratings of therapist empathy. Though it is substantively interesting that high- and low-empathy therapists clustered in particular settings, statistically, it would be ideal for site and empathy to be orthogonal. Given the number of sites and therapists in the present data, it is not possible to disentangle site and therapist influences.

<sup>2</sup> There were three coders in the study. Almost half of the included sessions were coded twice ( $n = 43$ , 48%). The remainder was coded once. Some sessions received scores below and above the high- and low-empathy cutpoints. In these cases, we used the mean of the ratings to determine classification in high- or low-empathy categories.

pitch (Juslin & Scherer, 2005), higher mean  $f_0$  scores are considered indicative of higher levels of encoded arousal. It is less expensive and complicated to capture than many other physiological measures (i.e., it can be extracted from an audio recording). Because therapist and SP were not recorded with separate microphones, audio files were manually decomposed (segmented) and stored in separate files that contained only therapist or SP speech. Mean  $f_0$  was estimated from segmented files every 0.25 s via the Praat speech signal processing software, with a bandpass filter of 75 to 350 Hz, which represents typical values for human speech (Boersma & Weenink, 2009).

**Statistical Analysis**

Several features of the current  $f_0$  data have implications for the statistical analyses. Most notably,  $f_0$  is an intensive longitudinal measure (taken every quarter second) that is highly imbalanced between speakers over time. An individual talk-turn could be less than a second (e.g., a typical back-channel like “uh-huh”) or as long as a minute or more. To model synchrony in therapist–SP mean  $f_0$  and its association with therapist empathy, we used multivariate multilevel models (Baldwin, Imel, Braithwaite, & Atkins, in press; MacCallum, Kim, & Malarkey, 1997), with therapist and SP mean  $f_0$  modeled as a bivariate outcome, nesting repeated observations of mean  $f_0$  within minutes and then within sessions. Multilevel models provide flexible approaches to modeling nested, longitudinal, and unbalanced data. To examine the association of therapist–SP mean  $f_0$ , we used the following model:

$$Y_{ijk} = b_0Therapist_k + b_1SP_k + b_2Therapist\ Gender_k + b_3Empathy_k + r_{Tj} + r_{SPj} + u_{Tk} + u_{SPk} + e_{ijk}, \quad (1)$$

where  $Y_{ijk}$  is the mean  $f_0$  estimate for observation  $i$  in minute  $j$  in session  $k$ . The overall intercept in regression models is absent in Equation 1, and two indicator variables,  $Therapist_{ijk}$  and  $SP_{ijk}$  are included, where  $Therapist_{ijk} = 1$  when the  $f_0$  estimate is from a therapist and 0 when it is from the SP, where  $SP_{ijk} = 1$  when the  $f_0$  estimate is from an SP and 0 when it is from a therapist. We examined differences in mean  $f_0$  between high- and low-empathy sessions with a dichotomous indicator ( $Empathy_k$ ). Another dichotomous indicator ( $Therapist.Gender_k$ ) adjusted for therapist gender (all SPs were female). The random effects,  $r_{Tj}$ ,  $r_{SPj}$  and  $u_{Tk}$ ,  $u_{SPk}$  modeled between-minute and between-session differences in mean  $f_0$  for therapists and SPs, respectively, and  $e_{ijk}$  is a residual error term. We assumed random effects were distributed multivariate normal:

$$\begin{bmatrix} u_{Tk} \\ u_{SPk} \\ r_{Tj} \\ r_{SPj} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_T}^2 & & & \\ \sigma_{u_{SP}}^2 & \sigma_{u_{SP}u_T} & & \\ & \sigma_{r_T}^2 & & \\ & \sigma_{r_{SP}r_T} & \sigma_{r_{SP}}^2 & \end{bmatrix} \right),$$

where  $\sigma_{r_T}^2$ ,  $\sigma_{r_{SP}}^2$ ,  $\sigma_{u_T}^2$ , and  $\sigma_{u_{SP}}^2$  are the variances in  $f_0$  estimates across therapists and SPs in minutes and sessions, respectively, whereas  $\sigma_{u_{SP}u_T}$  and  $\sigma_{r_{SP}r_T}$  are covariances between therapist and SP mean  $f_0$  for sessions and minutes, respectively. For interpretation, we transformed covariances to correlations. Finally, the re-

sidual error is assumed independent of the random effects and is normally distributed:

$$e_{ijk} \sim N(0, \sigma_e^2).$$

**Synchrony of SP and therapist mean  $f_0$ : Hypothesis 1.** The test of therapist–SP mean  $f_0$  synchrony was twofold. On the basis of the multivariate model shown in Equation 1, we examined the correlation of therapist and SP mean  $f_0$  across sessions and across minutes within each session. These correlations examined the degree of SP and therapist mean  $f_0$  similarity within minutes and sessions. As a sensitivity check, we randomly shuffled or “re-assigned” SPs to therapists from a different session (i.e., we matched mean  $f_0$  estimates of therapists and SPs that did not actually speak together in that session). Here, the correlation of therapist–SP mean  $f_0$  for both minute sessions should be close to zero as there was no opportunity for mutual influence (see Gurtman, 2001; Sadler, Ethier, Gunn, Duong, & Wood, 2009, for other examples of this approach).

**Association of empathy and mean  $f_0$  synchrony: Hypothesis 2.** To examine the association of empathy ratings with therapist–SP mean  $f_0$  synchrony, we fit a second multivariate multilevel model in which random effects were stratified by high- and low-empathy ratings (low empathy = 1 or 2, high empathy = 6 or 7). This model fits eight variance terms (SP-high empathy, therapist-high empathy, SP-low empathy, therapist-low empathy for both sessions and minutes) but allows covariances only within high- and low-empathy pairs. A larger association of therapist and SP mean  $f_0$  in minutes and sessions from sessions rated as high versus low empathy would suggest that mean  $f_0$  synchrony is associated with empathy ratings.

All analyses were done using Bayesian models fit via the MCMCglmm package in R (Hadfield, 2010; R Core Team, 2011). Bayesian analyses use an iterative Markov chain Monte Carlo (MCMC) fitting procedure that yields posterior distributions that are straightforward to use for interval estimation around all parameters. Minimally informative priors were used for both fixed effects (i.e., multivariate normal with large variances) and variance components (i.e., inverse Wishart; Gelman & Hill, 2007). Three MCMC simulations (or “chains”) were run for each model, including 100,000 iterations, a burn-in of 5,000 iterations, and a thinning interval of 20. Posterior convergence was assessed with autocorrelation and trace plots, as well as the Gelman-Rubin statistic. Fixed and random effects were estimated via the mode and 95% highest posterior density intervals (HPDs; Gelman & Hill, 2007), where an HPD interval is the Bayesian analogue of a classical confidence interval.

**Results**

Across the 89 sessions, there were 267,153 mean  $f_0$  observations ( $M = 3,001$  per session) and a total of 1,826 min ( $M = 20.5$  per session). The average mean  $f_0$  value across SPs and therapists was 145.2 ( $SD = 47.6$ ). All effects were adjusted for therapist gender ( $B = -45.7$ , 95% CI  $[-52.7, -36.4]$ ), as mean  $f_0$  is higher, on average, in women. We also adjusted for the particular SP who participated in the session. However, the HPD intervals for SP effects included zero ( $B_{sp1} = -11.3$ , 95% CI  $[-24.6, 0.84]$ ;

$B_{sp2} = 4.6$ , 95% CI  $[-5.7, 12.5]$ .<sup>3</sup> Mean  $f_0$  was lower in sessions coded as high empathy ( $B = -11.4$ , 95% HPD  $[-19.7, -3.7]$ ) compared with low-empathy sessions, suggesting that high-empathy sessions were characterized by somewhat lower levels of vocally expressed arousal. Figure 1 is a scatterplot of mean  $f_0$  for therapist and SP pairs across minutes and illustrates mean differences between high- and low-empathy sessions.

Results for Hypotheses 1 and 2 based on the multivariate multilevel models described earlier are shown in Figure 2. Correlations for session-level random effects were larger than minute-level correlations, but patterns of association were similar. Consistent with Hypothesis 1, random effects for therapist and SP mean  $f_0$  were correlated: sessions,  $r = .71$ , 95% HPD  $[.57, .80]$ ; minutes,  $r = .26$ , 95% HPD  $[.21, .32]$ . In addition, when therapists were randomly paired with SPs they did not treat, the correlation was close to zero: sessions,  $r = -.08$ , 95% HPD  $[-.31, .10]$ ; minutes,  $r = .02$ , 95% HPD  $[-.03, .09]$ . Consistent with Hypothesis 2, the correlation of therapist and SP mean  $f_0$  was higher in sessions with high-empathy ratings: sessions,  $r = .80$ , 95% HPD  $[.66, .88]$ ; minutes,  $r = .31$ , 95% HPD  $[.25, .39]$ , as compared with low-empathy ratings: sessions,  $r = .36$ , 95% HPD  $[-.03, .62]$ ; minutes,  $r = .20$ , 95% HPD  $[.10, .33]$ . As the model was fit using Bayesian methods, it is possible to derive a point estimate and interval of the difference between high- and low-empathy correlations using the posterior distributions (see Gelman & Hill, 2007, for a discussion of posterior distributions in MCMC analyses). The HPD interval for the difference between correlations from high- and low-empathy-rated sessions did not include zero: sessions,  $r = .39$ , 95% HPD  $[.16, .86]$ ; minutes,  $r = .11$ , 95% HPD  $[.01, .25]$ .

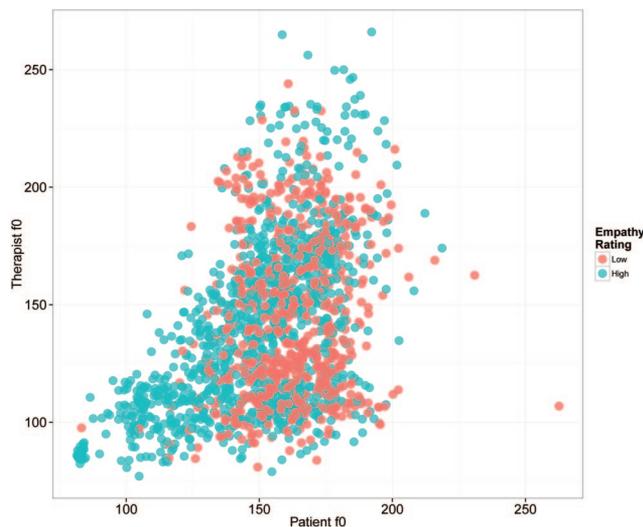


Figure 1. Scatterplot of the covariation of therapist and standardized patient mean fundamental frequency ( $f_0$ ) across 1,775 min (from 89 sessions). Red points are from sessions with low-empathy ratings, and blue points are from sessions with high ratings. There were 1,826 potential min—71 min did not include either an SP or therapist observation (all values in Hz). Points are semitransparent to show overlapping observations.

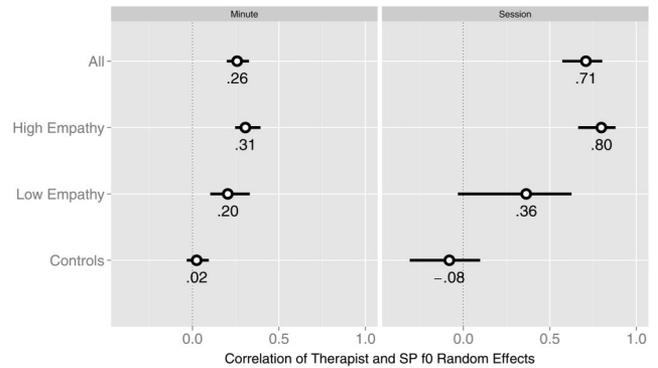


Figure 2. Posterior modes (circles) and 95% credible intervals for the correlation of standardized patient (SP) and therapist mean fundamental frequency random effects at minute and session level for (a) all sessions, (b) high-empathy sessions, (c) low-empathy sessions, and (d) controls; therapists were “digitally” matched with SPs they did not treat.

## Discussion

This study is the first in which the relationship of vocal acoustics to a measure of therapist performance in psychotherapy has been examined. The association of mean  $f_0$  extends prior evidence for synchrony in clinical interactions to a measure obtained from vocal recordings. The finding that mean  $f_0$  was higher in low-empathy sessions is consistent with research in couples therapy indicating that levels of  $f_0$  range during a pretherapy interaction predicted negative outcomes 2 years posttherapy (Baucom et al., 2009). In sum, the current study provides an important link between earlier studies of physiological arousal and psychotherapy (e.g., Marci et al., 2007), suggesting that synchrony in physiological arousal may be partially mediated through vocal cues such as mean  $f_0$ .

In contrast to most clinical trials that have a selected therapist sample with a restricted range of fidelity ratings, the current data included a large sample of therapists working in outpatient substance abuse treatment facilities. The use of a large therapist sample obtained in the *real world* improves the generalizability of findings to therapists working in the community. In addition, therapists were not selected for competence. Accordingly, empathy ratings spanned the full range of the scale. The use of three SPs reduced the impact of patient variance on the evaluation of therapists. For example, with only three SPs, the shuffling SP–therapist pairs occasionally resulted in instances of matching the same people (i.e., the same therapist and SP) but from different sessions. Specifically, it was possible for a therapist to be matched with the same SP they treated initially, but SP mean  $f_0$  values came from a different clinical interaction with that same therapist. The lack of synchrony in shuffled pairs indicates that mean  $f_0$  synchrony was likely a result of the interaction of therapist and SPs and was not due to a chance matching between therapist and SP vocal acoustics.

<sup>3</sup> Although Bayesian statistics do not yield a  $p$  value as is common in null hypothesis significance testing, HPD intervals that exclude zero have a broadly similar interpretation to 95% confidence intervals that exclude zero (i.e., significant at  $p < .05$ ).

## Limitations

The original data were collected as a part of a dissemination trial in which therapists were the primary focus of evaluation. Although SPs provide a common stimulus from which to evaluate therapists, it was not possible to connect observations of therapist behavior to future client outcomes. In addition, the use of SPs restricted the examination of how therapists respond to different patients. The use of only three SPs could also be seen as a limitation—a larger group could result in greater variability in mean  $f_0$  values and greater generalizability. Future studies might explore the performance of therapists with RPs. For example, it will be important to determine how measures of synchrony correlate with the client's experience of empathy.

A further limitation of this study includes the use of only one channel of conversational data. Specifically, the current study focused on the “how” channel (e.g., pitch/arousal) rather than the “what” channel (e.g., words). The actual words used by a therapist are likely to be important in the perception of empathy; thus, future work should also examine synchrony in word usage among therapists and patients. In addition, previous work indicates that the communication of empathy extends beyond conversational data alone and includes synchrony in nonverbal behaviors (Ramseyer & Tschacher, 2011) and physiology (e.g., skin conductance; Marci et al., 2007). However, this limitation would apply to all observer ratings of psychotherapy based on audio recordings, and the MITI is typically rated using audio only. In addition, measures of  $f_0$  are likely to be correlated with visual cues such as hardened facial expression or crossed arms (Scherer & Ellgring, 2007). Accordingly, linguistic methods would only miss aspects of nonverbal behavior that are unassociated with verbal cues. Finally, physiological measures can be complicated and invasive (Rochman & Amir, 2013) and do not provide direct tests of vocal mechanisms that may be involved in the communication of empathy.

## Clinical Implications

Although psychotherapy training places an emphasis on the specific words therapists use, vocal cues that are correlated with physiological arousal also appear to be clinically relevant. These cues may provide relationally significant information via both their absolute level and their covariation with another person. This is consistent with traditional counselor training efforts focused on the therapist *following* the client. Thus, the therapist should not only follow the clients' words but also track the extent to which they are in tune with clients' verbal tone.

The process by which vocal acoustics contribute to the experience and communication of empathy in MI and other treatments is an important area for future work in psychotherapy and doctor–patient communication generally. For example, mean  $f_0$  synchrony may be one mechanism by which the therapist generates an understanding of the client (e.g., it may promote the experience of empathy). Empathic therapists might experience a representation of a client's emotional state through perception of client vocal cues, such as mean  $f_0$ , and then intentionally or unintentionally use their internal experience to “guess” about how the client feels. Thus, mean  $f_0$  synchrony is merely the residue of a therapist working hard to understand the client, not a direct component of empathic communication. Alternatively, changes in vocally encoded arousal that result from attended perception may be a

mechanism by which empathy is communicated. Here, therapists are not simply using their internal state as another source of information to generate a well-timed reflection. Approximating the tone of the client is how therapists express empathy.

Another interesting area for future research may involve examining synchrony that is either *excessive* or associated with negative process or outcomes. There was evidence of therapist–SP synchrony even in sessions with low-empathy ratings. Synchrony in sessions with negative process indicators may be indicative of emotional contagion or a conflictual relationship characterized by increased arousal in both parties. As emotional contagion and empathy are thought to share common perception–action mechanisms, the distinction of the two constructs remains an important question (Iacoboni, 2009; Preston & De Waal, 2002).

Findings also eventually have relevance for the evaluation of MI in the community. Measures of empathy in MI and psychotherapy generally have relied on the *felt* sense of observers or participants. The gold standard for evaluating therapist behavior—observational coding—is laborious and often prone to bias. These limitations prevent large-scale efforts to evaluate the quality of treatments in the real world. Accordingly, it is not surprising that much of the raw data generated during psychotherapy goes unexamined and the structure of mechanisms involved in psychotherapy remain unclear (Webb, DeRubeis, & Barber, 2010). Novel indicators of therapist empathy that are low cost and unobtrusive (e.g., can be obtained from a microphone) such as mean  $f_0$  may lead to increases in coding efficiency and could have an important impact on public health through the development of clinically feasible methods for monitoring therapists.

## References

- Baer, J., Rosengren, D., Dunn, C., Wells, E., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug and Alcohol Dependence*, *73*, 99–106.
- Baer, J., Wells, E., Rosengren, D., Hartzler, B., Beadnell, B., & Dunn, C. (2009). Agency context and tailored training in technology transfer: A pilot evaluation of motivational interviewing training for community counselors. *Journal of Substance Treatment*, *37*, 191–202.
- Baldwin, S. A., Imel, Z. E., Braithwaite, S., & Atkins, D. (in press). Analyzing multiple outcomes using multivariate multilevel models. *Journal of Consulting and Clinical Psychology*.
- Baucom, B. R., Atkins, D. C., Simpson, L. E., & Christensen, A. (2009). Prediction of response to treatment in a randomized clinical trial of couple therapy: A 2-year follow-up. *Journal of Consulting and Clinical Psychology*, *77*, 160–173.
- Baucom, B. R., Saxbe, D., Ramos, M. R., Spies, L., Duman, S. R., & Margolin, G. (2012). Characteristics and correlates of adolescents' fundamental frequency during family conflict. *Emotion*, *12*, 1281–1291.
- Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (Version 5.1.05) [Computer program]. Retrieved from <http://www.praat.org/>
- Chartrand, T., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.
- Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 219–274). London, England: Elsevier.
- Diamond, G. M., Rochman, D., & Amir, O. (2010). Arousing primary vulnerable emotions in the context of unresolved anger: “Speaking

- about” versus “speaking to”. *Journal of Counseling Psychology*, 57, 402–410.
- Dimascio, A., Boyd, R. W., & Greenblatt, M. (1957). Physiological correlates of tension and antagonism during psychotherapy a study of “interpersonal physiology”. *Psychosomatic Medicine*, 19, 99–104.
- Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71–83). Cambridge, MA: MIT Press.
- Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy*, 48, 43–49.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multi-level/hierarchical models*. New York, NY: Cambridge University Press.
- Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Mahwah, NJ: Erlbaum.
- Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners’ speech. *Journal of Nonverbal Behavior*, 14, 237–251.
- Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research*, 11, 35–46.
- Gregory, S. W., Webster, S., & Huang, G. (1990). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication*, 13, 195–217.
- Gurtman, M. B. (2001). Interpersonal complementarity: Integrating interpersonal measurement with interpersonal models. *Journal of Counseling Psychology*, 48, 97–110.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY: Cambridge University Press.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.
- Imel, Z. E., Baldwin, S., Baer, J., Hartzler, B., Dunn, C., Rosengren, D., & Atkins, D. (2013). *Evaluating therapist adherence in motivational interviewing by comparing performance with standardized and real patients*. Manuscript submitted for publication.
- Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology*, 99, 549–571.
- Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22, 39–44.
- Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). New York, NY: Oxford University Press.
- Kappas, A., Hess, U., & Scherer, K. R. (1991). Voice and emotion. In R. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 200–238). New York, NY: Cambridge University Press.
- Lee, C. C., Katsamanis, A., Black, M. P., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2012). Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions. *Computer Speech & Language*. Advance online publication.
- Lee, C. C., Katsamanis, A., Black, M., Baucom, B., Georgiou, P. G., & Narayanan, S. S. (2011, August). *An analysis of PCA-based vocal entrainment measures in married couples’ affective spoken interactions*. Paper presented at INTERSPEECH 2011: 12th Annual Conference of the International Speech Communication Association. Florence, Italy.
- Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63, 234–246.
- MacCallum, R. C., Kim, C., & Malarkey, W. B. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32, 215–253.
- Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of Nervous and Mental Disease*, 195, 103–111.
- Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist*, 64, 527–537.
- Moyers, T. B., Martin, T., Catley, D., Harris, K. J., & Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: Reliability of the motivational interviewing skills code. *Behavioural and Cognitive Psychotherapy*, 31, 177–184.
- Moyers, T. B., Martin, T., & Manuel, J. (2005). *Motivational interviewing treatment integrity (MITI) coding system*. Retrieved from <http://casaa-0031.unm.edu>
- Moyers, T. B., Martin, T., Manuel, J. K., Hendrickson, S. M. L., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, 28, 19–26.
- Moyers, T. B., & Miller, W. R. (2013). Is low therapist empathy toxic? *Psychology of Addictive Behaviors*, 27, 878–884.
- Moyers, T. B., Miller, W. R., & Hendrickson, S. M. L. (2005). How does motivational interviewing work? Therapist interpersonal skill predicts client involvement in motivational interviewing sessions. *Journal of Consulting and Clinical Psychology*, 73, 590–598.
- Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21, 337–360.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169–190.
- Preston, S. D., & De Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1–20.
- Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, 79, 284–295.
- R Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Robinson, J. W. (1982). Autonomic responses correlate with counselor-client empathy. *Journal of Counseling Psychology*, 29, 195–198.
- Rochman, D., & Amir, O. (2013). Examining in-session expressions of emotions with speech/vocal acoustic measures: An introductory guide. *Psychotherapy Research*, 1–13.
- Rogers, C. (1980). *A way of being*. New York, NY: Houghton Mifflin.
- Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54, 329–349.
- Sadler, P., Ethier, N., Gunn, G. R., Duong, D., & Wood, E. (2009). Are we on the same wavelength? Interpersonal complementarity as shared cyclical patterns during interactions. *Journal of Personality and Social Psychology*, 97, 1005–1020.
- Scherer, K. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), *Handbook of psychophysiology: Emotion and social behavior* (pp. 165–197). London, England: Wiley.
- Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion*, 7, 158–171.
- Vader, A. M., Walters, S. T., Prabhu, G. C., Houck, J. M., & Field, C. A. (2010). The language of motivational interviewing and feedback: Counselor language, client language, and client drinking outcomes. *Psychology of Addictive Behaviors*, 24, 190–197.

- Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: The role of emotional similarity and empathic accuracy. *Emotion, 8*, 792–802.
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*, 200–211.
- Weusthoff, S., Baucom, B. R., & Hahlweg, K. (2013). Fundamental frequency during couple conflict: An analysis of physiological, behavioral, and sex-linked information encoded in vocal expression. *Journal of Family Psychology, 27*, 212–220.

Received May 15, 2013

Revision received September 4, 2013

Accepted September 5, 2013 ■