Received: 31 August 2021 | Revised: 19 October 2021 | Accepted: 21 October 2021

DOI: 10.1002/capr.12489

ORIGINAL ARTICLE

WILEY

It's not what you said, it's how you said it: An analysis of therapist vocal features during psychotherapy

Christina S. Soma[1] | Dillon Knox[2,3] | Timothy Greer[2] | Keith Gunnerson[1] | Alexander Young[2] | Shrikanth Narayanan[2]

[1]Department of Educational Psychology, University of Utah, Salt Lake City, Utah, USA

[2]Viterbi Department of Computer Science, University of Southern California, Los Angeles, California, USA

[3]Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California, USA

Correspondence
Christina Soma, Department of Educational Psychology, University of Utah, Salt Lake City, Utah, USA.
Email: tsoma15@gmail.com

Funding information
National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: R01/AA018673; National Institute on Drug Abuse, Grant/Award Number: R34/DA034860

Abstract

Psychotherapy is a conversation, whereby, at its foundation, many interventions are derived from the therapist talking. Research suggests that the voice can convey a variety of emotional and social information, and individuals may change their voice based on the context and content of the conversation (e.g. talking to a baby or delivering difficult news to patients with cancer). As such, therapists may adjust aspects of their voice throughout a therapy session depending on if they are beginning a therapy session and checking in with a client, conducting more therapeutic 'work' or ending the session. In this study, we modelled three vocal features—pitch, energy and rate—with linear and quadratic multilevel models to understand how therapists' vocal features change throughout a therapy session. We hypothesised that all three vocal features would be best fit with a quadratic function—starting high and more congruent with a conversational voice, decreasing during the middle portions of therapy where more therapeutic interventions were being administered, and increasing again at the end of the session. Results indicated a quadratic model for all three vocal features was superior in fitting the data, as compared to a linear model, suggesting that therapists begin and end therapy using a different style of voice than in the middle of a session.

KEYWORDS

emotional expression, linguistics, psychotherapy research, quantitative analysis, vocal acoustics

1 | INTRODUCTION

Psychotherapy is an effective mental health treatment whereby the majority of interventions are provided verbally (Lambert & Bergen, 2004) through a therapeutic conversation (Frank, 1961; Norcross, VandenBos, & Freedheim, 2011). 'Talk therapy'—originally coined by Freud as the 'talking cure' (Freud & Breuer, 1895)—describes the process of mental health care through which a therapist and client engage in a therapeutic conversation, with the goal of alleviating client distress. Broadly, effective psychotherapy involves the therapist and client making therapeutic goals, conducting therapeutic tasks to reach those goals, and the therapist conveying an understanding and compassion for the client's experiences (Barkham & Lambert, 2021; Bordin, 1979; Gelso, 2014). Individually, therapists learn a variety of theories, interventions and skills in order to provide scientifically grounded treatment to clients (e.g. evidence-based practice; Laska et al., 2014). For instance, cognitive therapy focuses on exploring the interactions with thoughts, feelings and behaviours

(Beck, 1979), whereas motivational interviewing (MI) relies on the therapist eliciting verbal indications of change from the client (i.e. change talk; Miller & Rollnick, 2012). However, even within the same treatment paradigm, there are variations within therapists on the delivery of interventions (Beutler et al., 2004). One such variation could be how the therapist modulates their voice—or a 'therapy voice'.

Generally, vocal acoustics convey a variety of affective information (see Juslin & Scherer, 2005, for a review) and help individuals to engage in effective social–emotional communication (Lima et al., 2013). Studying vocal acoustics in psychotherapy is a quickly developing research area (e.g. Weusthoff et al., 2018) that has provided important insights into therapeutic processes such as emotional expression (Baucom et al., 2012; Rochman & Amir, 2013), emotion coregulation (Bryan et al., 2018; Soma et al., 2020; Wieder & Wiltshire, 2020), emotional synchrony (Gaume et al., 2019; Imel et al., 2014), client responses to interventions (e.g. higher pitch during empty chair exercise; Diamond et al., 2010), diagnostically relevant symptom severity (e.g. slower speaking rate for clients with depression; Mundt et al., 2012) and correlates to treatment outcomes (e.g. Baucom et al., 2009). Verbal expression in psychotherapy has been captured through a variety of vocal features (see Rochman & Amir, 2013 for a review), such as fundamental frequency, or *f0* (Weusthoff et al., 2013; Yang et al., 2012), speaking rate (e.g. Diamond et al., 2010; Ornston et al., 1968), and volume or intensity (e.g. Knowlton & Larkin, 2006). These studies help to describe complex therapeutic processes occurring between the therapist and client and demonstrate that studying vocal acoustics is an appropriate psychotherapy research methodology.

Components of a therapist's voice may contribute to the therapeutic process whereby the therapist uses their voice to communicate emotions, attempt to soothe the client or convey understanding of the client's experience (Bady, 1985). Defining and assessing therapist voices has been done at a high level, characterising therapists' voices, for instance, as warm (Morris & Magrath, 1979; Morris & Suckerman, 1974), cold (Morris & Suckerman, 1974) and productive (Wiseman & Rice, 1989). Introductory counselling skills textbooks often encourage beginning therapists to be warm or friendly to convey empathy or interest (e.g. Nelson-Jones, 2002; Sommers-Flanagan & Sommers-Flanagan, 2018). Researchers have also demonstrated that human-rated factors such as warmth, interest and curiosity may positively contribute to building a strong therapeutic alliance (Ackerman & Hillsenroth, 2003). One study compared muscle relaxation treatment for a specific phobia with either a 'warm' or 'cold' therapist voice, and a no-treatment control. Warm voices were characterised as 'soft, melodic and pleasant', and cold voices as 'harsh, impersonal and businesslike' (Morris & Suckerman, 1974, p. 245). Results indicated that participants receiving snake phobia treatment from therapists using the warm-voice protocol improved significantly over the other two treatments (Morris & Suckerman, 1974). Though providing initial evidence that therapist vocal style changes could contribute to the therapeutic process, there is a lack of understanding of the specific vocal features (e.g. pitch, volume, speaking rate) and how these features change throughout a session.

Implications for Practice and Policy

- Therapists deliver the majority of their interventions verbally. However, little is known about how therapists change their voice during therapy sessions.
- Therapist vocal changes may correspond to changes in session content and helping therapists understand how the features of their voice contribute to the psychotherapy process may be of value.
- This research could help therapists understand the impact of their voice and potentially contribute to training and supervision of additional therapeutic variables that impact the therapeutic process.
- Knowledge about therapist vocal features may additionally aid licensed therapists to continue to understand their impact on clients and potentially advance their practice

Researchers have posited that combining several vocal features may provide a better picture of vocal expression (Bone et al., 2014; Chaspari et al., 2017). Several studies have investigated a combination of vocal features to understand how therapists might create a therapeutically advantageous voice. In a study investigating how therapists convey empathy, researchers demonstrated that higher vocal pitch and energy (a proximal measure of loudness) corresponded to lower perceptions of therapist empathy (Xiao et al., 2014). In another study, Knowlton and Larkin (2006) compared treatment outcomes for therapists using a 'recommended therapy voice' or 'conversational voice' during progressive relaxation training (see Bernstein & Borkovec, 1973, for original protocol). The recommended therapy voice was indicated by a decrease in the pitch, volume and rate of speaking, and the conversational voice was given no specific instructions on how to alter their voice besides to speak 'conversationally'. Results demonstrated that the treatment group receiving therapy from the recommended voice had a significant reduction in distress symptoms. Though postulating that particular vocal features may be beneficial during therapy, these studies are lacking specificity of the moments when therapists project particular features and how these features might fluctuate throughout the session.

To our knowledge, the extant psychotherapy literature lacks further study on how therapists' vocal features change throughout a session. However, outside of psychotherapy, research on how physicians, surgeons and nurses deliver information to patients has demonstrated some context-dependent differences in vocal features used by healthcare providers (Haskard et al., 2008). For example, researchers audio-recorded conversations between oncologists and their patients, and noted times when the oncologist was delivering bad news to the patient and when the oncologist was hosting more typical medical conversation (e.g. scheduling, discussing symptoms) during an appointment. Results demonstrated that nearly all providers reduced their speaking rate and pitch when

delivering bad news. The authors hypothesised that the intention of these vocal changes was used to convey care or sympathy to their patients (McHenry et al., 2012). In another study, participants rated segments of a provider–patient interaction with different characteristics (e.g. warm–cold, submissive–dominant), and researchers conducted an analysis of the vocal features present during each category. Results demonstrated that when the provider was perceived as being more dominant, there was an increase in volume and speaking rate (Harrigan et al., 1989). However, the study lacks information about when this occurred during the session. Unfortunately, much like the psychotherapy literature, research on specificity of vocal features and modelling when vocal changes happen in other healthcare interactions is sparse.

On the contrary, researchers exploring other intimate relationships outside of the healthcare field have found evidence of changes in vocal features during particular times in conversation, which may serve specific purposes. For instance, the way parents speak to their infants is a well-established research area (i.e. infant-directed speech, motherese, or 'baby talk'; see Fernald & Kuhl, 1987; Golinkoff et al., 2015), and the implications of studying how parents change their voices can be critical for understanding the development of language acquisition skills. It has been shown that parents increase the range of their fundamental frequency (*f0*) up to 800Hz (as opposed to the typical 300Hz when addressing adults), decrease their speaking rate and adjust certain vowel sounds when speaking to attract or hold their infant's attention (Fernald et al., 1989; Fernald & Kuhl, 1987). Infants of parents who struggle with matching some of these vocal features due to mood or affect issues may show more difficulties with language acquisition (e.g. Kaplan et al., 2002). Outside of the parent–child relationship, linguistics researchers posit there are a multitude of contextual and environmental factors that influence how and when individuals speak to each other (Heylighen & Dewaele, 1999). For example, second-language teachers articulate their vowel sounds more clearly than those engaged in casual conversation in order to avoid misunderstanding and promote an ease of language learning (Eddine, 2011). Sports commentators and auctioneers will drastically change their voices—using increases in amplitude and speaking rate—to fill expectations in their particular occupational role (Kuiper, 1995). Much like when individuals must provide specific information or are trying to attract others' attention or appear more likeable, psychotherapists may use changes in their voices during different portions of therapy to facilitate a specific therapeutic process.

A central component of therapy is the therapist engaging the client in as much therapeutic conversation as possible (i.e. 'work'; Hill & Stephany, 1990; Horvath & Bedi, 2002; Kozart, 2002). However, in addition to the 'work' portion of therapy, there may be an initial check-in (Dobson & Dobson, 2013; Hill, 2009) or some small talk (Nelson-Jones, 2002), which is not necessarily meant to be therapeutic, but begins the session. For example, in a standard cognitive behavioural therapy (CBT) session of 50 min, Dobson and Dobson (2013; see also Beck, 2011) recommend 5–10 min to be spent on a check-in, gauging client distress, brief discussions of the prior week, and setting a schedule for the session. Small talk in healthcare settings may largely serve as a bonding mechanism between the provider and patient, and acts as a conversational transition to more medically relevant topics (for a review, see Benwell & McCreaddie, 2016; Defibaugh, 2018). Though psychotherapy research is largely lacking research on the function of small talk, it is likely that small talk or the 'check-in' portion of a therapy session may serve a similar function—that is, a transition into more therapeutically oriented conversation.

Dobson and Dobson (2013) posit that the next 30–40 min of a session constitutes the 'working' phase, which includes discussing thoughts, emotions and behaviours (see also Beck, 1979). Therapeutic work across treatments may also include discussing the therapeutic relationship (Horvath & Bedi, 2002; Norcross, 2010), interpersonal process between the therapist and client (e.g. Henry et al., 1990; Sullivan, 1968), and how therapy is progressing (e.g. symptom outcomes; see Lambert et al., 2001). Researchers have attempted to identify moment-to-moment sequences of therapist behaviours through coding systems (e.g. verbal response modes [Stiles, 1979]; Psychodynamic Intervention Rating Scale [Milbrath et al., 1999]; Therapeutic Collaboration Coding System [Ribeiro et al., 2013]). Additionally, researchers have used the linguistic technique of conversational analysis (Sacks et al., 1974) in an attempt to sequence therapeutic conversation and identify patterns of a therapist's interventions and a client's response (Gobbo, 2010; Peräkylä, 2019; Voutilainen et al., 2011). In a study utilising a conversational analysis framework, researchers categorised the words spoken by the therapist and client throughout the session, and then coded sequences of words that appeared throughout the session. Coding was conducted with the Therapeutic Cycles Model (Mergenthaler, 1996), and the coding system indicated instances where the researchers considered therapeutic work to be happening (e.g. emotion processing, discussing problems). Results demonstrated a specific sequence of events that had significantly higher therapeutic language and emotional tone, indicating therapeutic work towards the middle of the therapy session (Lepper & Mergenthaler, 2007). The work portion of a therapy session may also be indicated by specific speech patterns, not just the words spoken. For instance, in a study of non-verbal behaviours, Hill and Stephany (1990) asked therapists and clients to watch and rate behaviours that occurred during a therapy session. When observing when therapeutic work was happening, therapists had significantly more speech hesitations during times the therapist perceived the client to be engaged in therapeutic work. Though the study, as well as the broader literature, lacks an understanding of fluctuations in other vocal features, there may be evidence of therapists changing the way they talk to clients throughout the session.

The final 5–10 min of Dobson and Dobson's recommended session structure (2013) is summarising and planning homework. For CBT, homework is considered an essential way to end the session (see Beck, 2011). However, different treatment paradigms may not necessarily assign homework or follow-up exercises to the client, and have other methods of ending a session. For example, therapists may note the time as a method for ending the session—for example 'our time is up' (for a review, see Gans, 2016), or have a particular phrase

that begins to summarise or indicates a session end—for example 'I'll see you next week' (Brody, 2009; Gabbard, 1982). Researchers in the broader medical field have also discussed the phenomenon of 'doorknob syndrome', whereby the patient discloses a medically relevant detail in the last moments of an appointment. Physicians may prepare for this moment by asking questions that indicate closure—for example, 'Is there anything else bothering you?' (Jackson, 2005). However, much like the check-in or small talk portion of therapy, the extant literature has not explored how vocal features may change from these portions of therapy to the work portion of therapy.

In sum, there is a developing research area investigating therapists' vocal acoustics to understand therapeutic processes; however, this research is lacking exploration of the moment-to-moment changes for a combination of vocal features that make up how the therapist conveys therapeutic content—that is, a therapy voice. Medical and linguistics researchers have provided some basis for why individuals would change their voice based on the person, context and content of the conversation. As such, the current study aims to model how three specific vocal features—pitch, energy and speaking rate (see Juslin & Scherer, 2005; Knowlton & Larkin, 2006)—are changing over the course of a psychotherapy session. We hypothesised that therapists will have more of a conversational voice, indicated by a higher pitch, volume and speaking rate towards the beginning portion of therapy (e.g. small talk or check-in), a 'therapy voice' during the work portions of therapy, indicated by a decrease in pitch, volume and rate, and return to the conversational voice towards the end of therapy. We modelled each of the three vocal features with both linear and quadratic multilevel models, hypothesising that a convex quadratic model will provide the best fit for the data.

2 | METHODS

2.1 | Data

We analysed the vocal features from a sample of 212 psychotherapy sessions[1] obtained from six motivational interviewing (MI) dissemination trials (Baer et al., 2009; Lee et al., 2013, 2014; Neighbors et al., 2012; Roy-Byrne et al., 2014; Tollison et al., 2008). Data from 104 therapists were utilised, all of whom conducted single-session therapy with at least one client (n = 61; M = 2.04, SD = 2.04), with one therapist seeing a maximum of 12 clients. The trials included therapists who completed training and received weekly supervision (Lee et al., 2013, 2014; Neighbors et al., 2012; Tollison et al., 2008), therapists who received initial training and were notified of drift from MI protocol (Roy-Byrne et al., 2014), and therapists who received training without continued supervision (Baer et al., 2009). In one study, primary care providers conducted either brief alcohol and drug interventions or enhanced care (Roy-Byrne et al., 2014). In three of the studies, providers targeted alcohol and marijuana use in college students (Lee et al., 2014; Neighbors et al., 2012; Tollison et al., 2008). In the last study, providers were conducting therapy in community-based primary care clinics, in which patients may have been using many types of drugs at one time (Baer et al., 2009). The treatment modality in each study was MI, an evidenced-based treatment that emphasises empathy and specifies that the therapist uses a specific type of language (e.g. reflective statements), focused on eliciting verbal statements regarding change in a behaviour from clients (Miller & Rollnick, 2012). The trials were conducted in the Pacific Northwest and were approved through the University of Washington IRB.

Participants in all clinical trials were promised to have no identifying information outside of the content of the recorded session. As such, there is no demographic information for the sample. However, according to the applicable census at the time of data collection (2000), five million Washington state residents reported their race as White (84.9%), Black or African American (4%), Asian (6.7%), American Indian and Alaska Native (2.7%), Native Hawaiian and Other Pacific Islander (0.7%), or indicated another race (4.9%); in a separate measure, 7.5% indicated Hispanic or Latino ethnicity (Evans et al., 2001). Census data collected gender identification only for the gender binary (i.e. male and female) and did not include space for data on non-binary and transgender individuals. Given this limitation, the data reported that 50.2% of Washington residents identified as female and 49.8% as male. Participants across studies were 18 years and older, either in a primary care setting or on a college campus.

2.2 | Vocal features

We analysed three primary vocal features that have been hypothesised to affect the therapeutic process: pitch, voice volume and speaking rate (see Knowlton & Larkin, 2006). Volume was estimated with a proxy measure—energy—a well-validated measure (Bone et al., 2014). Pitch and energy were extracted with the *Kaldi* toolkit (Povey et al., 2011), using a frame size of 25ms, with a 10-ms gap between measurements. Utterance-level representations for each feature were computed by taking the mean of all frames of a given utterance. Speaking rate was calculated at the utterance level by estimating the number of syllables that occurred in a given utterance, and then dividing by the total duration of that utterance. To estimate syllables, each utterance was broken up into individual transcribed words, and *syllapy* (Jacobs & Kinder, 2020) was then used to estimate the number of syllables for each. The final utterance-level speaking rate was then calculated from these two measures. Note that speaking rate, as used here, includes any pauses and non-vocalised sounds in the duration calculation (see Arnfield et al., 1995, for more discussion).

2.3 | Session timing

The total session time ranged from 7.67 to 81.5 min (M = 39.81; SD = 18.6). Given the large range of real-world clinical interactions,

we chose to analyse all sessions and to represent time as a per cent of the session that has occurred. We scaled each time point within a given session measured in seconds to a per cent of how much of the session had occurred. Thus, each time point was represented on the scale of [0,1].

2.4 | Analysis

2.4.1 | MLM Models

We used multilevel modelling (MLM; Raudenbush & Bryk, 2002; Singer et al., 2003) to determine how the changes in therapist vocal features varied as a function of session length. Each of the three vocal features—pitch, energy and rate—was modelled separately and predicted by the portion of total session time that had occurred. For comparison, therapist vocal features were modelled with both linear and quadratic MLMs, where observations of vocal features (level 2) are nested within sessions (level 1). We hypothesised that the quadratic models would indicate a superior fit to the data. Regression analysis was conducted using the *lme4* package (Bates et al., 2014) in the *R* programming language (version 3.5.2; R Development Core Team, 2019), using full maximum-likelihood estimation procedures; the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and log likelihood (LL) were inspected for each model and informed model selection. We then compared linear and quadratic model fit with a chi-squared test (χ^2). The equation for each feature is as follows, with Y_{ij} representing each predicted vocal feature.

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * Proportion + r_{ij}$$

Level 2 model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

3 | RESULTS

Refer to Table 1 for coefficients, 95% confidence intervals, and fit indices for linear and quadratic models of each vocal feature. Figure 1 visually represents the predicted values based on the selected superior models. The figure was generated with the *ggplot2* package (Wickham et al., 2019). Each plot contains a smoothed series of values from the raw data computed with the *geom_smooth* function, which utilises a LOESS smoother to demonstrate the general data

TABLE 1 Multilevel model comparison of therapist vocal features

Vocal Feature	Model	β_{ij}	Residual Variance (Std. Dev.)	95% CI	Model Fit Index		
					AIC	BIC	LL
Pitch	Linear		0.99 (0.99)		102,960.7	102,994.7	−51,476.4
		β_{00}= 0.03*		[0.01, 0.05]			
		β_{10}= −0.05**		[−0.09, −0.02]			
	Quadratic		0.99 (0.99)		102,919.6	102,962.1	−51,454.8
		β_{00}= 0.10***		[0.07, 0.13]			
		β_{10}= −0.51***		[−0.65, −0.37]			
		β_{20}= 0.45***		[0.32, 0.59]			
Energy	Linear		0.99 (0.99)		102,854	102,888	−51,423
		β_{00}= 0.10***		[0.08, 0.11]			
		β_{10}= −0.19***		[−0.23, −0.16]			
	Quadratic		0.99 (0.99)		102,794.0	102,836.5	−51,392.0
		β_{00}= 0.19***		[0.16, 0.22]			
		β_{10}= −0.74***		[−0.88, −0.60]			
		β_{20}= 0.54***		[0.41, 0.68]			
Rate	Linear		0.99 (0.99)				
		β_{00}= 0.04***		[0.02, 0.06]	102,948	102,982	−51,470
		β_{10}= −0.08***		[−0.12, −0.04]			
	Quadratic		0.99 (0.99)				
		β_{00}= 0.08***		[0.05, 0.11]	102,939.5	102,982.0	−51,464.7
		β_{10}= −0.31***		[−0.45, −0.17]			
		β_{20}= 0.22**		[0.09, 0.36]			

Note: Significant level at *$p < .05$; **$p < .01$ and ***$p < .001$.

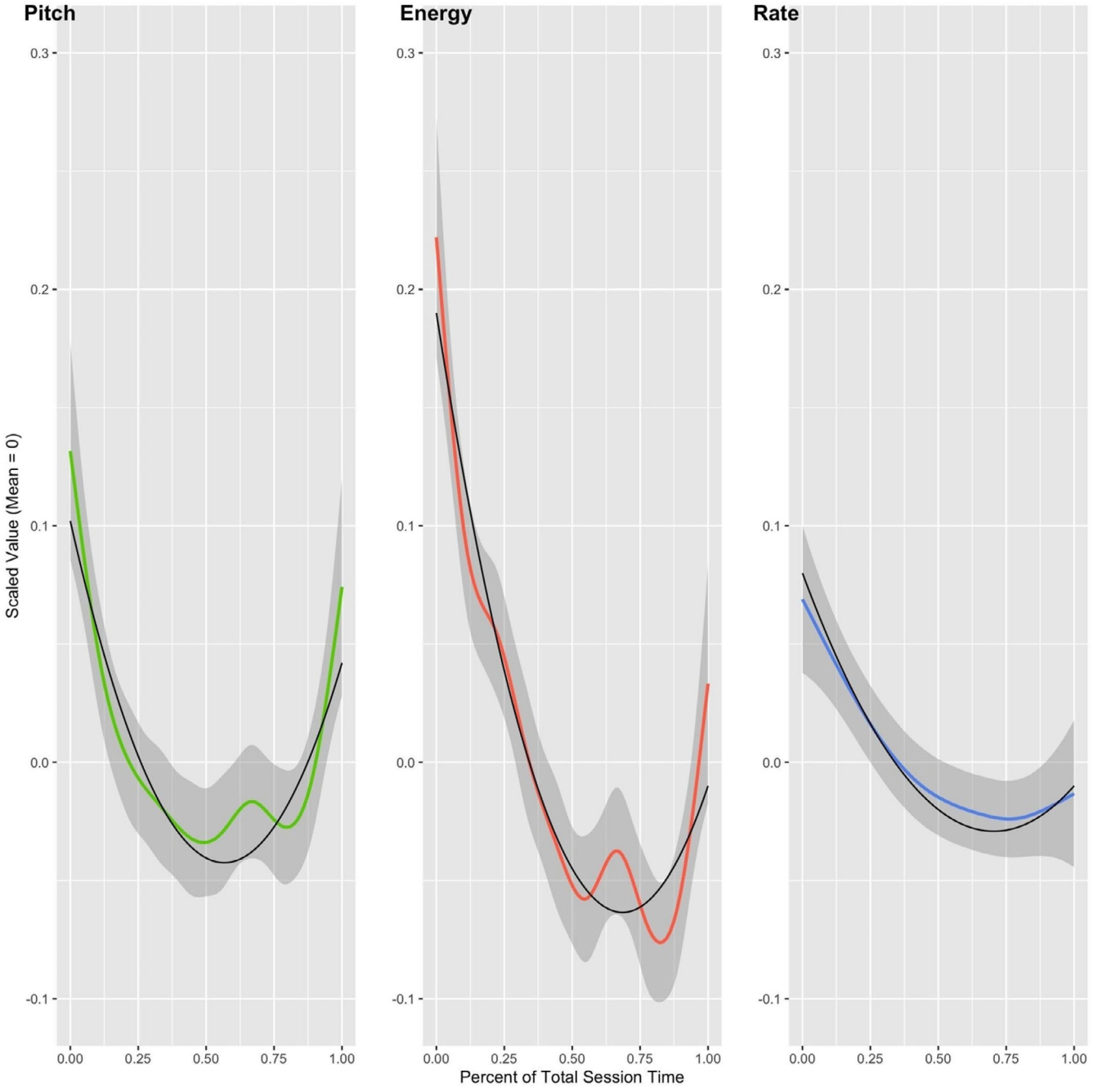

FIGURE 1 Comparison of Smoothed Raw Data and Model Predicted Data for Three Vocal Features. Note: The colour line with error area is the raw data; the black line is the model predicted features

trend (Wickham & Sievert, 2016). Each plot also contains the predicted values from the superior fitting model.

3.1 | Pitch

The model fit indices indicated that the quadratic model fit for pitch was superior, with the AIC_{Quad} and BIC_{Quad} indicating lower values, and the LL_{Quad} indicating a higher value. The linear (β_{10}= −0.51) and quadratic (β_{20} = 0.50) coefficients were all significant, p <.01.

The quadratic model performed significantly better than the linear model, $\chi^2(1) = 43.11$, p <.001, indicating that the therapists' pitch started at a higher value, decreased until a minimum point, and then increased. When the therapist's voice changed from negative to positive, slope in the quadratic equation (i.e. the critical point or relative minimum) was −0.04. The relative minimum occurred after 55.78% of the session had occurred. The values where the therapist's voice changed from positive to negative (i.e. x-intercepts) were 0.26 and then back to positive at 0.85. That is, the therapist's pitch was below their session-level average after 26% of the session had

occurred and then above their average after 85% of a session had occurred.

3.2 | Energy

Similar to pitch, the model fit indices indicated that the quadratic model fit was superior for energy, with the AIC_{Quad} and BIC_{Quad} indicating lower values, and the LL_{Quad} indicating a higher value. The quadratic model performed significantly better than the linear model, $\chi^2(1) = 62.03$, $p < .001$, indicating again that the therapists' energy started high, dipped to a minimum and then increased again. The relative minimum predicted by the model was −0.06, indicating the point at which the therapist's energy was no longer decreasing, and beginning to increase, occurring after 67.72% had occurred. The first x-intercept was 0.34 where the therapist's energy started to be below the mean, and the second was projected to be 1.03; however, this value would be after the session has ended and thus is not interpretable. That is, the therapist's energy was below the mean for the remainder of the session.

3.3 | Speaking rate

The model fit indices indicated that the quadratic model fit was superior, with the AIC_{Quad} and BIC_{Quad} still indicating slightly lower values, and the LL_{Quad} indicating a slightly higher value. The quadratic model performed significantly better than the linear model, $\chi^2(1) = 10.47$, $p < .01$, indicating that the therapists' speaking rate started high, dipped lower and again increased. The relative minimum predicted by the model was −0.03, indicating the point at which the therapist's speaking rate is no longer decreasing, and begins to increase, occurring after 68.52% of the session. The first x-intercept was 0.34 where the therapist's speaking rate values dipped below their average, and the second was projected to be 1.07; however, this value would also occur after the session ended and thus, similar to energy, is not interpretable. The therapist's speaking rate stayed below their mean for the remainder of the session.

4 | DISCUSSION

Prior research has demonstrated that people change their voices within and across specific contexts, ranging from adults using 'baby talk' with infants (e.g. Fernald & Kuhl, 1987) to an oncologist delivering potentially difficult news to a patient during a medical visit (McHenry et al., 2012). However, investigations into how a therapist's voice might change throughout therapy are sparse. This is one of the first to model how a therapist's voice—specifically the pitch, energy and speaking rate—changes throughout a psychotherapy session. Psychotherapy theory and pedagogy has explored the overarching structure of therapy broadly containing a check-in or some small talk at the beginning, therapeutic work during the middle, and a way of indicating that the session is coming to an end. Prior research has indicated that therapists may generally use a voice specific for therapy with the potential to improve treatment outcomes (Knowlton & Larkin, 2006), as well as showing specific changes to their speech patterns, such as more hesitations during the work portion of therapy (Hill & Stephany, 1990). Therapists may also indicate the end of a therapy session with a particular phrase or topic (Gans, 2016). As such, we hypothesised that the therapists' vocal features we measured would start higher, indicating a more conversational or casual style of talking, decrease and reach their lowest point (see Knowlton & Larkin, 2006), and then increase towards the end of the session. Statistically, we hypothesised this pattern would be best fit by a quadratic function. Using multilevel modelling, we concluded that the trajectory of change for the therapists' pitch, energy and rate was all best fit by quadratic models. These findings confirmed our hypothesis that a quadratic model would best fit pitch and energy, and described the trajectory of pitch and energy changing from high to low, and returning to higher values.

In particular, the therapists' pitch steadily decreased until about half of the session had occurred, at which point the therapists began to increase the pitch of their voice until the end of the session. Similar to pitch, the therapists' energy trajectory was best predicted by a quadratic model, with the therapists' energy changing from decreasing to increasing after about two thirds of the session had occurred. The quadratic model for speaking rate was a significantly better fit than the linear model; however, the model fit indicators showed only slight changes between the models. Similar to energy, the therapists' speaking rate changed from decreasing to increasing also at about two thirds of the way through the session. Taken together, the changes in these three features show a fluctuating 'therapy voice', which demonstrates some consistencies with prior literature regarding the structure and content of therapy sessions.

In the aforementioned study from Knowlton and Larkin (2006), the recommended therapy voice had participants speak with a lower pitch, energy and speaking rate, whereas the conversational voice had higher pitch, energy and speaking rates. Though the study did not explore how the vocal feature values were changing throughout sessions, the distinction between the therapy and conversational voice may provide context for why the therapists' vocal features in our study fluctuated from high to low and back to high again. First, the higher values of pitch, energy and speaking rate at the beginning of a session could potentially indicate a more conversational bridge to start therapy. Small talk during medical visits has been posited to have many roles, including both helping patients acclimate to more difficult medical conversations and to distract the patient during a visit (Benwell & McCreaddie, 2016). Research with patient–surgeon dyads has indicated that there may even be a specific transition from small talk to the focused medical content after the patient has included some context and background information about their visit (Hudak & Maynard, 2011). The therapist may be utilising a more conversational voice to begin the session, provide the client with an opportunity to ease into the session, and provide the therapist with information that may lead to more therapeutic content.

As the therapy dyads transition into the middle portions of the session, the pitch, energy and speaking rate decrease, more indicative of Knowlton and Larkin's recommended therapy voice, perhaps occurring in conjunction with therapeutic work that is happening. Using the critical points from each quadratic model, our results indicated that the therapists' vocal features were at their lowest between half and two thirds of the way through the session, which is consistent with research showing that therapeutic topics (e.g. emotion processing) are discussed significantly more in the middle of a session (e.g. Lepper & Mergenthaler, 2007). Perhaps in conjunction with therapeutic topics being discussed during the middle of the therapy session, therapists are lowering their pitch, energy and speaking rate (e.g. speech hesitations; Hill & Stephany, 1990) to facilitate therapeutic conversations. Prior research has also indicated that therapists' pitch and energy negatively correlate with human-rated session-level empathy (Xiao et al., 2014). Perhaps therapists lower their pitch and energy to convey empathy during portions of therapy that may be more indicative of therapeutic work.

For each of the three features studied, the model indicated an increase towards the end of the session. Theorists have posited that some therapists may have a particular phrase, style or indicator they utilise in order to indicate that the session is coming to a close or to wrap up the session (Brody, 2009; Gabbard, 1982). In a study that utilised conversational analysis to study 69 medical encounters of individuals with their surgeons, results demonstrated that small talk often occurred at the end of a session, transitioning patient–surgeon dyads to close the visit (Jin, 2018). Perhaps similar small talk vocal features trends are occurring in both the beginning and ending of therapy sessions, thereby transitioning the dyad in and out of therapeutic work. Though there was a trend among all features, there were some differences where each feature started and ended. In particular, energy started at a higher level than where it ended at the end of session, whereas pitch and rate did not demonstrate the same degree of difference between the beginning and end. It is possible that higher therapist energy at the beginning of a session may correspond to the therapist attempting to show their engagement in the therapeutic process, therefore demonstrating openness to client experiences (i.e. therapeutic presence), possibly promoting the client's perception of safety (Geller & Porges, 2014) and starting to engage the client in the therapeutic process. As the session progresses, however, the therapist seems to become less emotionally activated, as evidenced by the lower energy, to provide the client space (Levitt, 2001). Towards the end of the session, the client may be more emotionally activated, and thus, the therapist may act to ground, or regulate, the client, and the therapist could do this by speaking more softly and slowly (see Paz et al., 2021; Soma et al., 2020).

In addition to the general quadratic trend of each vocal feature, the critical value where the therapist's voice changed from decreasing to increasing, the model's x-intercepts, provides the points at which the therapist's vocal features are above or below their session-level average. Pitch, energy and speaking rate transitioned from above to below the session-level mean after about 30% of the session had progressed, consistent with Dobson and Dobson's (2013) proposed structure of a cognitive therapy session. Though Figure 1 demonstrates the values for pitch, energy and speaking rate have a relatively small range, even small deviations from the therapist's average speech could be meaningful. The therapist may not necessarily want to significantly change their vocal features, and the client may be able to discern small changes in their voice. For example, in a study of the just noticeable difference (JND) phenomenon—originated from Ernst Weber—researchers demonstrated individuals can discern a difference in speaking rate of just 5% (Quené, 2007). Therapists may be able to convey emotional expression, communicate interventions and show their interest in the client without deviating their vocal features significantly away from the mean.

4.1 | Limitations and future directions

The study presents promising findings on which future studies may build; however, there are some limitations. The data do not include demographic features to explore dynamics of changes in vocal features across cultures, race and ethnicity, and gender expression. Future research may investigate how the voice may fluctuate when therapy dyads differ (e.g. White therapists and clients of colour versus White therapists with White clients), and how client perceptions of the therapist's attention to cultural variables and the presence of cultural conversations may correspond to fluctuations in vocal features. In addition, the data include single-session therapy encounters. Future studies may explore how these changes in vocal features evolve over time both with experience and across different clients, as well as how such patterns may correspond to client outcomes. Further, direct comparisons between how the therapist is speaking during a therapy session to their speech patterns outside of therapy would provide a deeper understanding of how the therapist may use and change their voice as part of the therapy process. Though this study does not directly compare therapists in and outside the therapy room, we provide a foundation for asking more complex questions about how the voice is a tool and part of therapist interventions and overall intention of where to guide the session. The implications of our research may inform future pedagogy and how therapists in training use their voice as part of how they engage with clients.

ORCID

Christina S. Soma https://orcid.org/0000-0002-3607-6643

ENDNOTE

[1] The full data set included real and standardised patients. We included only real patient sessions in the sample.

REFERENCES

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review*, *23*(1), 1–33. https://doi.org/10.1016/S0272-7358(02)00146-0

Arnfield, S., Roach, P., Setter, J., Greasley, P., & Horton, D. (1995). Emotional stress and speech tempo variation. In Speech under Stress

Bady, S. L. (1985). The voice as a curative factor in psychotherapy. *Psychoanalytic Review*, *72*(3), 479–490.

Baer, J. S., Wells, E. A., Rosengren, D. B., 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 Abuse Treatment*, *37*(2), 191–202. https://doi.org/10.1016/j.jsat.2009.01.003

Barkham, M., & Lambert, M. J. (2021). The efficacy and effectiveness of psychological therapies. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (7th ed., in press). Wiley.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

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. https://doi.org/10.1037/a0014405

Baucom, B. R., Saxbe, D. E., Ramos, M. C., Spies, L. A., Iturralde, E., Duman, S., & Margolin, G. (2012). Correlates and characteristics of adolescents' encoded emotional arousal during family conflict. *Emotion*, *12*(6), 1281. https://doi.org/10.1037/a0028872

Beck, A. T., (Ed.). (1979). *Cognitive therapy of depression*. Guilford press.

Beck, J. (2011). *Cognitive behavior therapy: Basics and beyond*, 2nd ed. Guilford Press.

Benwell, B., & McCreaddie, M. (2016). Keeping "small talk" small in health-care encounters: Negotiating the boundaries between on- and off-task talk. *Research on Language and Social Interaction*, *49*(3), 258–271. https://doi.org/10.1080/08351813.2016.1196548

Bernstein, D. A., & Borkovec, T. D. (1973). *Progressive relaxation training: A manual for the helping professions*. Research Press.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 27–307). John Wiley & Sons.

Bone, D., Lee, C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. S. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1162–1177. https://doi.org/10.1044/2014_JSLHR-S-13-0062

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252–260. https://doi.org/10.1037/h0085885

Brody, S. (2009). On the edge: Exploring the end of the analytic hour. *Psychoanalytic Dialogues*, *19*(1), 87–97. https://doi.org/10.1080/10481880802634560

Bryan, C. J., Baucom, B. R., Crenshaw, A. O., Imel, Z., Atkins, D. C., Clemans, T. A., Leeson, B., Burch, T. S., Mintz, J., & Rudd, M. D. (2018). Associations of patient-rated emotional bond and vocally encoded emotional arousal among clinicians and acutely suicidal military personnel. *Journal of Consulting and Clinical Psychology*, *86*(4), 372–383. https://doi.org/10.1037/ccp0000295

Chaspari, T., Timmons, A. C., Baucom, B. R., Perrone, L., Baucom, K. J., Georgiou, P., Margolin, G., & Narayanan, S. S. (2017, October). *Exploring sparse representation measures of physiological synchrony for romantic couples*. In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 267–272). IEEE.

Defibaugh, S. (2018). Caring as competent: Small talk in medical visits. *Nurse practitioners and the performance of professional competency* (pp. 101–120). Palgrave Macmillan.

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*(4), 402. https://doi.org/10.1037/a0021115

Dobson, D., & Dobson, K. (2013). In-Session Structure and Collaborative Empiricism. *Cognitive and Behavioral Practice*, *20*(4), 410–418. https://doi.org/10.1016/j.cbpra.2012.11.002

Eddine, A. N. (2011). *Second language acquisition: The articulation of vowels and the importance of tools in the learning process*. The Acquisition of L2 Phonology, 1.

Evans, D., Price, J., & Barron, W. (2001). *Profiles of general demographic characteristics: 2000 Census of population and housing*. Washington, DC: US Department of Commerce.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510. https://doi.org/10.2307/1130938

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Frank, J. D. (1961). *Persuasion and Healing*. The Johns Hopkins University Press.

Freud, S., & Breuer, J. (1895). *Studies on hysteria* (pp. 255–305). Hogarth.

Gabbard, G. O. (1982). The exit line: Heightened transference-countertransference manifestations at the end of the hour. *Journal of the American Psychoanalytic Association*, *30*(3), 579–598. https://doi.org/10.1177/000306518203000302

Gans, J. S. (2016). " Our Time is Up": A Relational Perspective on the Ending of a Single Psychotherapy Session. *American Journal of Psychotherapy*, *70*(4), 413–427. https://doi.org/10.1176/appi.psychotherapy.2016.70.4.413

Gaume, J., Hallgren, K. A., Clair, C., Schmid Mast, M., Carrard, V., & Atkins, D. C. (2019). Modeling empathy as synchrony in clinician and patient vocally encoded emotional arousal: A failure to replicate. *Journal of Counseling Psychology*, *66*(3), 341–350. https://doi.org/10.1037/cou0000322

Geller, S. M., & Porges, S. W. (2014). Therapeutic presence: Neurophysiology mechanisms mediate feeling safe in therapeutic relationships. *Journal of Psychotherapy Integration*, *24*(3), 178–192.

Gelso, C. (2014). A tripartite model of the therapeutic relationship: Theory, research, and practice. *Psychotherapy Research*, *24*(2), 117–131. https://doi.org/10.1080/10503307.2013.845920

Gobbo, F. (2010). CoCAL–Constructive Conversation Analysis: Applications to Therapeutic Settings (White Paper).

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Harrigan, J. A., Gramata, J. F., Lucic, K. S., & Margolis, C. (1989). It's how you say it: Physicians' vocal behavior. *Social Science & Medicine*, *28*(1), 87–92. https://doi.org/10.1016/0277-9536(89)90310-9

Haskard, K. B., Williams, S. L., DiMatteo, M. R., Heritage, J., & Rosenthal, R. (2008). The provider's voice: Patient satisfaction and the content-filtered speech of nurses and physicians in primary medical care. *Journal of Nonverbal Behavior*, *32*(1), 1–20. https://doi.org/10.1007/s10919-007-0038-2

Henry, W. P., Schacht, T. E., & Strupp, H. H. (1990). Patient and therapist introject, interpersonal process, and differential psychotherapy outcome. *Journal of Consulting and Clinical Psychology*, *58*(6), 768–774. https://doi.org/10.1037/0022-006X.58.6.768

Heylighen, F., & Dewaele, J. M. (1999). Formality of language: Definition, measurement and behavioral determinants. *Interner Bericht, Center "Ieo Apostel", Vrije Universiteit Brüssel*, *4*.

Hill, C. E. (2009). *Helping skills: Facilitating, exploration, insight, and action*. American Psychological Association.

Hill, C. E., & Stephany, A. (1990). Relation of nonverbal behavior to client reactions. *Journal of Counseling Psychology*, *37*(1), 22–26. https://doi.org/10.1037/0022-0167.37.1.22

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work. Therapist contributions and responsiveness to patients* (pp. 37–70). Oxford University Press.

Hudak, P. L., & Maynard, D. W. (2011). An interactional approach to conceptualising small talk in medical interactions. *Sociology of Health & Illness*, *33*(4), 634–653. https://doi.org/10.1111/j.1467-9566.2011.01343.x

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology*, *61*(1), 146–153. https://doi.org/10.1037/a0034943

Jackson, G. (2005). "Oh… by the way…": doorknob syndrome.

Jacobs, A. M., & Kinder, A. (2020). Quasi error-free text classification and authorship recognition in a large corpus of english literature based on a novel feature set. *arXiv preprint arXiv:2010.10801*.

Jin, Y. (2018). Small talk in medical conversations: Data from China. *Journal of Pragmatics*, *134*, 31–44. https://doi.org/10.1016/j.pragma.2018.06.011

Juslin, P. N., & Scherer, K. R. (2005). *Vocal expression of affect*. Oxford University Press.

Kaplan, P. S., Bachorowski, J. A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science*, *13*(3), 268–271. https://doi.org/10.1111/1467-9280.00449

Knowlton, G. E., & Larkin, K. T. (2006). The influence of voice volume, pitch, and speech rate on progressive relaxation training: Application of methods from speech pathology and audiology. *Applied Psychophysiology and Biofeedback*, *31*(2), 173–185. https://doi.org/10.1007/s10484-006-9014-6

Kozart, M. F. (2002). Understanding efficacy in psychotherapy: An ethnomethodological perspective on the therapeutic alliance. *American Journal of Orthopsychiatry*, *72*, 217–231. https://doi.org/10.1037/0002-9432.72.2.217

Kuiper, K. (1995). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Routledge.

Lambert, M. J., & Bergen, G. (2004). Overview, Trends and Future Issues. In M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. John Wiley & Sons.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, *69*(2), 159–172. https://doi.org/10.1037/0022-006X.69.2.159

Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy*, *51*(4), 467. https://doi.org/10.1037/a0034332

Lee, C. M., Kilmer, J. R., Nieghbors, C., Atkins, D. C., Zheng, C., Walker, D. D., & Larimer, M. E. (2013). Indicated Prevention for College Student Marijuana Use: A Randomized Controlled Trial. *Journal of Counseling and Clinical Psychology*, *81*(4), 702–709. https://doi.org/10.1037/a0033285

Lee, C. M., Nieghbors, C., Lewis, M. A., Kaysen, D., Mittman, A., Geisner, I. M., Atkins, D. C., Zheng, C., Garberson, L. A., Kilmer, J. R., & Larimer, M. E. (2014). Randomized Controlled Trial of a Spring Break Intervention to Reduce High-Risk Drinking. *Journal of Counseling and Clinical Psychology*, *82*(2), 189–201. https://doi.org/10.1037/a0035743

Lepper, G., & Mergenthaler, E. (2007). Therapeutic collaboration: How does it work? *Psychotherapy Research*, *17*(5), 576–587. https://doi.org/10.1080/10503300601140002

Levitt, H. M. (2001). Sounds of silence in psychotherapy: The categorization of clients' pauses. *Psychotherapy Research*, *11*(3), 295–309. https://doi.org/10.1080/713663985

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245. https://doi.org/10.3758/s13428-013-0324-3

McHenry, M., Parker, P. A., Baile, W. F., & Lenzi, R. (2012). Voice analysis during bad news discussion in oncology: Reduced pitch, decreased speaking rate, and nonverbal communication of empathy. *Supportive Care in Cancer*, *20*(5), 1073–1078. https://doi.org/10.1007/s00520-011-1187-8

Mergenthaler, E. (1996). Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, *64*, 1306–1315. https://doi.org/10.1037/0022-006X.64.6.1306

Milbrath, C., Bond, M., Cooper, S., Znoj, H. J., Horowitz, M. J., & Perry, J. C. (1999). Sequential consequences of therapists' interventions. *The Journal of Psychotherapy Practice and Research*, *8*(1), 40.

Miller, W. R., & Rollnick, S. (2012). *Motivational Interviewing: Helping People Change*. Guilford Press.

Morris, R. J., & Magrath, K. H. (1979). Contribution of therapist warmth to the contact desensitization treatment of acrophobia. *Journal of Consulting and Clinical Psychology*, *47*(4), 786. https://doi.org/10.1037/0022-006X.47.4.786

Morris, R. J., & Suckerman, K. R. (1974). Therapist warmth as a factor in automated systematic desensitization. *Journal of Consulting and Clinical Psychology*, *42*(2), 244. https://doi.org/10.1037/h0036261

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Neighbors, C., Lee, C. M., Atkins, D. C., Lewis, M. A., Kaysen, D., Mittmann, A., Fossos, N., Geisner, I. M., Zheng, C., & Larimer, M. E. (2012). A randomized controlled trial of event-specific prevention strategies for reducing problematic drinking associated with 21st birthday celebrations. *Journal of Consulting and Clinical Psychology*, *80*(5), 850–862. https://doi.org/10.1037/a0029480

Nelson-Jones, R. (2002). *Essential counselling and therapy skills: The skilled client model*. Sage.

Norcross, J. C. (2010). The therapeutic relationship. In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (pp. 113–141). American Psychological Association.

Norcross, J. C., VandenBos, G. R., & Freedheim, D. K. (2011). History of Psychotherapy: Continuity and Change. *American Psychological Association*.

Ornston, P. S., Cicchetti, D. V., Levine, J., & Fierman, L. B. (1968). Some parameters of verbal behavior that reliably differentiate novice from experienced psychotherapists. *Journal of Abnormal Psychology*, *73*(3p1), 240. https://doi.org/10.1037/h0025862

Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Laufer-Goldshtein, B., Narayanan, S., Keshet, J., & Atzil-Slonim, D. (2021). Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. *Journal of Consulting and Clinical Psychology*, *89*(3), 227. https://doi.org/10.1037/ccp0000623

Peräkylä, A. (2019). Conversation analysis and psychotherapy: Identifying transformative sequences. *Research on Language and Social Interaction*, *52*(3), 257–280. https://doi.org/10.1080/08351813.2019.1631044

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G., & Vesel, K. (2011). *The Kaldi speech recognition toolkit*. In IEEE 2011 workshop on automatic speech recognition and understanding (No. CONF). IEEE Signal Processing Society

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, *35*(3), 353–362. https://doi.org/10.1016/j.wocn.2006.09.001

R Development Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (Vol. 1)*. Sage Publications.

Ribeiro, E., Ribeiro, A. P., Gonçalves, M. M., Horvath, A. O., & Stiles, W. B. (2013). How collaboration in therapy becomes therapeutic: The therapeutic collaboration coding system. *Psychology and Psychotherapy: Theory, Research and Practice*, *86*(3), 294–314. https://doi.org/10.1111/j.2044-8341.2012.02066.x

Rochman, D., & Amir, O. (2013). Examining in-session expressions of emotions with speech/vocal acoustic measures: An introductory guide. *Psychotherapy Research*, *23*(4), 381–393. https://doi.org/10.1080/10503307.2013.784421

Roy-Byrne, P., Bumgardner, K., Krupski, A., Dunn, C., Ries, R., Donovan, D., West, I. I., Maynard, C., Atkins, D. C., Graves, M. C., Joesch, J. M., & Zarkin, G. A. (2014). Brief intervention for problem drug use in safety-net primary care settings: A randomized clinical trial. *JAMA*, *312*(5), 492–501. https://doi.org/10.1001/jama.2014.7860

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. In J. N. Schenkein (Ed.), *Studies in the organization of conversational interaction*. Academic Press.

Singer, J. D., Willett, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Soma, C. S., Baucom, B. R., Xiao, B., Butner, J. E., Hilpert, P., Narayanan, S., Atkins, D. C., & Imel, Z. E. (2020). Coregulation of therapist and client emotion during psychotherapy. *Psychotherapy Research*, *30*(5), 591–603. https://doi.org/10.1080/10503307.2019.1661541

Sommers-Flanagan, J., & Sommers-Flanagan, R. (2018). *Counseling and psychotherapy theories in context and practice: Skills, strategies, and techniques*. John Wiley & Sons.

Stiles, W. B. (1979). Verbal response modes and psychotherapeutic technique. *Psychiatry*, *42*(1), 49–62. https://doi.org/10.1080/00332747.1979.11024006

Sullivan, H. S. (1968). *Theory of interpersonal relations*. McGraw-HillBook Company.

Tollison, S. J., Lee, C. M., Neighbors, C., Neil, T. A., Olson, N. D., & Larimer, M. E. (2008). Questions and reflections: The use of motivational interviewing microskills in a peer-led brief alcohol intervention for college students. *Behavior Therapy*, *39*(2), 183–194. https://doi.org/10.1016/j.beth.2007.07.001

Voutilainen, L., Peräkylä, A., & Ruusuvuori, J. (2011). Therapeutic change in interaction: Conversation analysis of a transforming sequence. *Psychotherapy Research*, *21*(3), 348–365. https://doi.org/10.1080/10503307.2011.573509

Weusthoff, S., Baucom, B. R., & Hahlweg, K. (2013). The siren song of vocal fundamental frequency for romantic relationships. *Frontiers in Psychology*, *4*, 439. https://doi.org/10.3389/fpsyg.2013.00439

Weusthoff, S., Gaut, G., Steyvers, M., Atkins, D. C., Hahlweg, K., Hogan, J., Zimmermann, T., Fischer, M. S., Baucom, D. H., Georgiou, P., Narayanan, S., & Baucom, B. R. (2018). The Language of Interpersonal Interaction: An Interdisciplinary Approach to Assessing and Processing Vocal and Speech Data. *The European Journal of Counselling Psychology*, *7*(1), 69–85. https://doi.org/10.5964/ejcop.v7i1.82

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Sievert, C. (2016). *Ggplot2: Elegant graphics for data analysis*. ProQuest Ebook Central. https://ebookcentral-proquest-com.ezproxy.lib.utah.edu

Wieder, G., & Wiltshire, T. J. (2020). Investigating coregulation of emotional arousal during exposure-based CBT using vocal encoding and actor–partner interdependence models. *Journal of Counseling Psychology*, *67*(3), 337–348. https://doi.org/10.1037/cou0000405

Wiseman, H., & Rice, L. N. (1989). Sequential analyses of therapist-client interaction during change events: A task-focused approach. *Journal of Consulting and Clinical Psychology*, *57*(2), 281–286. https://doi.org/10.1037/0022-006X.57.2.281

Xiao, B., Bone, D., Segbroeck, M. V., Imel, Z. E., Atkins, D. C., Georgiou, P. G., & Narayanan, S. S. (2014). *Modeling therapist empathy through prosody in drug addiction counseling*. In Fifteenth Annual Conference of the International Speech Communication Association.

Yang, Y., Fairbairn, C., & Cohn, J. (2012). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*, 142–150. https://doi.org/10.1109/T-AFFC.2012.38

AUTHOR BIOGRAPHIES

Christina S. Soma is a PhD candidate at the University of Utah, working with Dr. Zac Imel. She recently returned from a research fellowship at the Modum Bad Psychiatric Hospital in Vikersund, Norway. She researches how the little moments in psychotherapy contribute to the broader therapist-client process. She is currently a doctoral psychology intern at the Colorado State University Health Network.

Dillon Knox is a PhD student under the supervision of Dr. Shrikanth Narayanan in the Department of Electrical Engineering at the University of Southern California. His research interests include music information retrieval and audio feature engineering. He most recently completed his master's degree in Electrical Engineering from USC.

Timothy Greer is a PhD candidate with the Signal Analysis and Interpretation Lab in the Department of Computer Science at the University of Southern California. Prior to USC, he worked at MIT's Lincoln Laboratory on natural language processing and graph analytics. Timothy's main research interests include multimodal music processing, media understanding and affective computing.

Keith Gunnerson completed his master's degree in Clinical Mental Health Counseling and more recently completed his PhD in Counseling Psychology while working in the Lab for Psychotherapy Science. His research interests include religiosity, therapeutic processes and first-generation college students. He is currently completing his postdoctoral placement in the Georgia Southern University Counseling Center.

Alexander Young most recently completed his bachelor's degree in Computer Science from the Viterbi School of Engineering at the University of Southern California. He is currently completing a preparatory phase for pursuit of a PhD in Computer Science at Universität des Saarlandes. His research interests revolve primarily around computational understanding of humour but also extend to all human–computer interaction.

Shrikanth Narayanan is a university professor and Niki & C. L. Max Nikias Chair in Engineering at the University of Southern California, and holds appointments as Professor of Electrical and Computer Engineering, Computer Science, Linguistics, Psychology, Neuroscience, Otolaryngology-Head and Neck Surgery, and Pediatrics, Research Director of the Information Science Institute, and a director of the Ming Hsieh Institute. His research focuses on developing engineering approaches to understand the human condition and in creating machine intelligence technologies that can support and enhance human experiences. He is a fellow of the Acoustical Society of America, IEEE, ISCA, the American Association for the Advancement of Science, the Association for Psychological Science, the American Institute for Medical and Biological Engineering and the National Academy of Inventors.

How to cite this article: Soma, C. S., Knox, D., Greer, T., Gunnerson, K., Young, A., & Narayanan, S. (2021). It's not what you said, it's how you said it: An analysis of therapist vocal features during psychotherapy. *Counselling and Psychotherapy Research*, 00, 1–12. https://doi.org/10.1002/capr.12489