Speech Rhythm Auto-Recurrence is Negatively Linked to Quality of Mental-Health Counseling Interactions

Vera Kempe[1] (v.kempe@abertay.ac.uk), Mark A. Adams[2] (m.a.adams@reading.ac.uk),
George Lovell[1] (p.g.lovell@abertay.ac.uk), Janet McLean[1] (j.mclean@abertay.ac.uk),
Kate Smith[1] (kate.smith@abertay.ac.uk)

[1]Division of Psychology, Abertay University,
Bell Street, Dundee DD1 1HG, UK

[2]School of Psychology & Clinical Language Sciences,
Harry Pitt Building, University of Reading, Reading GR6 6AL, UK

Abstract

We explored use of Recurrence Quantification Analysis (RQA) of speech rhythm data from mental-health counseling sessions for prediction of quality of psychotherapy. Time-series of inter-syllable intervals (ISIs) were extracted from 239 counseling sessions conducted by 12 therapists who repeatedly interacted with 30 clients. We found a negative association between recurrence metrics and client-rated session quality and a negative link between percent of laminarity and therapist-rated session quality, after controlling for self-reported client depression and distress measures and duration of speech sound within a session. Placing value on reduced recurrence in patterns of ISIs, and especially reduced degree of a dyadic system remaining in the same speech-rhythm pattern may be indicative of a desire for variation in content and strategies of client-therapist interaction. These exploratory findings point to the possibility of RQA-based automated systems to capture the 'footprint' of the non-verbal dynamic that is indicative of successful mental-health counseling.

Keywords: Recurrence Quantification Analysis, non-verbal speech parameters, speech rhythm, dialogue, psychotherapy

Introduction

Increased demand for mental health services has created a need for automated evaluations of psychotherapy to improve training and provide quality control in an expanding market. However, the cost- and time-prohibitive nature of coding and analyzing large amounts of linguistic data by human raters precludes timely evaluation of counseling sessions. A possible solution where psychotherapy rating tools use natural-language processing (e.g., Femotomos, Martinez, Chen, Singla et al., 2021) can run into privacy problems as therapists and clients may be reluctant to disclose the content of their interactions even to automated systems. In this study, we explore whether counseling efficacy can be estimated from the dynamics of non-verbal speech parameters. Exploring how well the dynamics of non-verbal aspects of communication can predict attainment of counseling goals is not just of applied importance but has important implications for the theoretical understanding of the structural organization of social interaction across multiple levels of linguistic representation and modalities of action.

Previous research has demonstrated that degree of alignment and coordination is linked to communicative success, especially when it occurs on more global levels of interaction (Fusaroli, Bahrami, Olsen, Roepstorff et al., 2012) or over longer time frames (Reitter & Moore, 2014). In laboratory experiments, the link between alignment and coordination on the one hand and communicative success on the other has been observed across multiple levels of linguistic representation, e.g., in lexical choices, prosody and speech rhythm (Fusaroli & Tylén, 2016). Yet, while solving joint tasks tends to benefit from alignment and coordination on various aspects of communication, in more naturalistic interactions such as psychotherapy the link to communicative outcomes is less straightforward: Several studies have shown positive links between synchronization of head and body movement and therapeutic outcomes (Paulick, Deisenhofer, Ramseyer, Tschacher et al., 2018; Ramseyer & Tschacher, 2011; 2014) while another study demonstrated that greater pitch synchrony was associated with lower ratings of therapeutic alliance and greater client distress, presumably reflecting sensitivity to therapist-initiated bids for changes in conversational direction or to attempts at repairing ruptures in rapport (Reich, Berman, Dale & Levitt, 2014). Thus, patterns of alignment and coordination between interlocutors may differ considerably depending on communicative situations and goals.

The aim of the present study was therefore to investigate further how patterns of alignment and coordination between interlocutors are linked to outcome measures in the specific setting of psychotherapeutic counseling. We examine links between clients' state of mental health, which is known to impact a range of communication features (Alpert, Pouget & Silva, 2001), dynamic characteristics of alignment on a non-verbal speech parameter – speech rhythm – and subjective evaluations of the quality of counselling sessions. We selected speech rhythm as the target non-verbal speech parameter due to its demonstrated suitability as a predictor of communicative outcome (Fusaroli & Tylén, 2016, Reuzel, Embregts, Bosman, Cox et al., 2013) and used RQA to quantify the temporal dynamics of speech rhythm.

Recurrence Quantification Analysis (RQA)

Time series data of features of communicative interaction tend to be non-stationary, i.e., the means and standard deviations of relevant measures change over time. Such

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

complex non-linear dynamic, which reflects shifting patterns of alignment and coordination over time, can be analyzed using RQA, a method for visualizing the patterns in which time series revisit previous states in reconstructed multidimensional state space using two-dimensional recurrence plots (Webber & Zbilut, 2005). RQA captures dynamic properties of communication that otherwise would be missed using aggregate measures or conventional time-series analyses. Characteristics of recurrence plots can be quantified using a range of recurrence metrics that can be subjected to statistical analyses (see Table 1).

Table 1: Recurrence metrics

Recurrence Parameter	Description
%Recurrence	Percent of recurrence points falling within a certain radius in state space
%Determinism (DET)	Percent of recurrence points forming diagonal lines; quantifies recurrence of deterministic patterns/sequences
Length (L)	Average length of diagonal lines; quantifies degree of stability and structure in a system
Entropy (ENTR)	Entropy of lengths of diagonal lines; quantifies complexity of recurrent patterns
%Laminarity (LAM)	Percent of recurring points falling into vertical/horizontal lines; indicates degree of system being trapped in the same state
Trapping Time (TT)	Average length of vertical/horizontal lines; indicates length of time a system being trapped in the same state

Many studies of alignment and coordination in dyadic interactions use cross-RQA (Coco & Dale, 2014) to analyze the way in which two or more interacting systems revisit each other's states. However, in this study we used auto-RQA which treats the dyad as a single system without discriminating between interlocutors. We chose this approach for two reasons: (a) Fusaroli & Tylén (2016) showed that auto-recurrence, which captures synergistic patterns of how interlocutors complement rather than mimic each other, is a better predictor of joint action success, and (b) size and nature of our corpus rendered diarization of audio-recordings to partition the speech stream by speaker identity technically infeasible. For communication that accompanies joint action the degree and complexity of recurrence obtained from auto-RQA should be positively correlated with communicative success and action outcome. However, it is unclear whether such a positive link would also be observed in psychotherapy where therapists aim not just to project empathy and to gain rapport but also to challenge and modify a client's entrenched views, behaviors, and communication strategies.

Method

We analyzed auto-recurrence patterns in audio-recordings of a corpus of counseling sessions conducted in the context of therapist training within the Pluralistic Counseling Framework (Cooper & McLeod, 2011; Smith & De La Prida, 2021). The aim was to establish whether RQA metrics can serve as predictors for subjective ratings of session quality by therapists and clients, controlling for clients' state of mental health and overall duration of speech sounds within a session.

The Counseling Corpus

The corpus was recorded as part of a community counseling research project between 2015 and 2018. All therapists and clients had provided consent for their data to be used in future research. The entire corpus comprises 644 sessions conducted by 17 therapists with 45 clients. The number of interactions per dyad ranged from 2 to 33 sessions. For the present analyses, were selected the first 6-9 sessions of each dyad (excluding the introductory session) from a sub-sample of 239 sessions of dyads for which more than six sessions were available. This corresponds to a counseling period that is roughly in line with therapeutic guidelines for individuals with mild-to-moderate depressive symptoms (National Institute for Health and Care Excellence [NICE], Guideline 1.5.3.6., 2009). Our sub-corpus comprised 30 clients (aged 21 to 65 years; 21 women, 9 men) who interacted with 12 therapists (11 women, 1 man). Five of the therapists interacted with one client, three interacted with two clients, the remaining therapists interacted with three, four, five or seven clients each. Counseling sessions lasted from 7 to 89 minutes with a mean session duration of 50 minutes and a standard deviation of 10 minutes.

Extracting Speech Duration and Rhythm

We used an algorithm written for PRAAT (Boersma & Weenink, 2018) by De Jong and Wempe (2009) to identify voiced intensity peaks as proxies for vowel onsets that identify individual syllables. We then extracted the time-series of intervals between these voiced intensity peaks which we take as a proxy for inter-syllable-intervals (ISIs). Note that these time-series are not a measure of speech rate in voiced segments of the interaction but rather a measure of overall speech rhythm that includes pauses and longer periods of silence.

To obtain the overall duration of speech sound, corresponding to the amount of talk within a session, we subtracted the sum of all ISIs from the total session duration.

Measuring of Client Mental Health

A battery of mental health questionnaires was administered prior to each session. We controlled for measures of client depression and distress because of their established link to communicative outcomes (Ellgring, 2007). This allows us to ascertain whether speech rhythm recurrence patterns account for additional variance in the appraisal of session quality over and above a clients' state of mental health.

Measuring Distress Client distress was evaluated using the Clinical Outcomes in Routine Evaluation (CORE) instrument, designed to determine treatment response during a course of psychotherapy with respect to a broad spectrum of problems associated with mental health difficulties including well-being deficits, mental health symptoms, life-functioning difficulties, and risk of harm (Evans, Mellor-Clark, Margison, Barkham et al., 2000). Total scores range from 0 to 136; the mean score in this sample was 46.1, SD = 18.1. CORE scores were missing for 13 of the 239 sessions.

Measuring Depression Client depression was evaluated using the Patient Health Questionnaire (PHQ-9), a monitoring instrument comprising multiple-choice questions designed to screen for severity of depressive symptoms in the general population (Kroenke, Spitzer, & Williams, 2001). Total scores range from 0 to 27 (scores of 5–9 are classified as mild depression; 10–14 as moderate depression; 15–19 as moderately severe depression; ≥ 20 as severe depression); the mean score in this sample was 10.6, SD = 5.5. PHQ scores were missing for 2 of the 239 sessions.

Session Quality Ratings

After each session both client and therapist rated the session for Helpfulness (1-9), Merit (1-7) and Productivity (1-7) with respect to their therapeutic goals. Cronbach's α was .86 for therapists and .77 for clients. Ratings for Merit and Productivity were rescaled and combined with the ratings for Helpfulness into an average score of session quality. Therapists' ratings were missing for 8, and clients' ratings were missing for 24 of the 239 sessions.

On average, therapists rated session quality lower than clients (client M = 6.81, SD = 1.15; therapist M = 5.44, SD = 1.03); this difference was significant in a multi-level model nesting session within clients within therapists ($\beta = -1.38$, $t = -9.59$, $p < .001$). Both ratings were positively linked in a multi-level regression model that predicted therapists' ratings from clients' ratings ($\beta = 0.52$, $t = 12.23$, $p < .001$).

Results

For all analyses, we fitted multiple multi-level regression models of session-level data with random effects of sessions nested within clients nested within therapists using the *lme4* package version 1.1.27.1 (Bates, Mächler, Bolker, & Walker, 2015) in R version 4.1.1. All fixed effects were centered and statistical significance of each model coefficient was evaluated with p-values approximated with the Satterthwaite method implemented in the *lmerTest* R-package version 3.1.3. (Kuznetsova, Brockhoff & Christensen, 2017).

Recurrence metrics were determined using the *crqa* package (Coco & Dale, 2014). To reconstruct the phase-space using the method of time-delayed embedding (Takens, Rand & Young, 1981), time delay for all sessions was estimated using the first local minimum of the *average-mutual-information* function which was 1 for all sessions as would be expected for inter-event intervals (Wallot & Leonardi, 2018). Embedding dimension was estimated using the *false nearest neighbors* algorithm (Abarbanel, 1996), yielding a range of 3 - 18 embedding dimensions across all sessions. We used an embedding dimension of 10 for all sessions as 96% of session embedding dimensions were at or below this value and RQA-metrics tend to be robust across a range of embedding dimensions (Wallot & Leonardi, 2018).

Because the sessions differed greatly in length using the same radius led to large variation in %Recurrence, which in most instances exceeded the recommended range of 1-5% (Webber & Zbilut, 2005) and rendered recurrence patterns indiscriminable in the recurrence plots. We therefore adjusted radius for each session to yield a %Recurrence within the range of 2.0 - 2.002%. Limitations in computational processing capacity made it necessary to cap the ISI time-series for 27 sessions at 11,000 data points. To control for recurrence due to incidental artefacts in the distribution of ISIs we also determined the recurrence metrics for time series of values shuffled within a session.

Figure 1 illustrates the variability in recurrence plots by depicting example plots of sessions with highest and lowest values in pre-session distress scores and post-session outcome quality ratings. Plot brightness reflects differences in amount of data points due to differences in amount of talk and suggest that higher levels of distress were associated with overall lower amounts of talking.

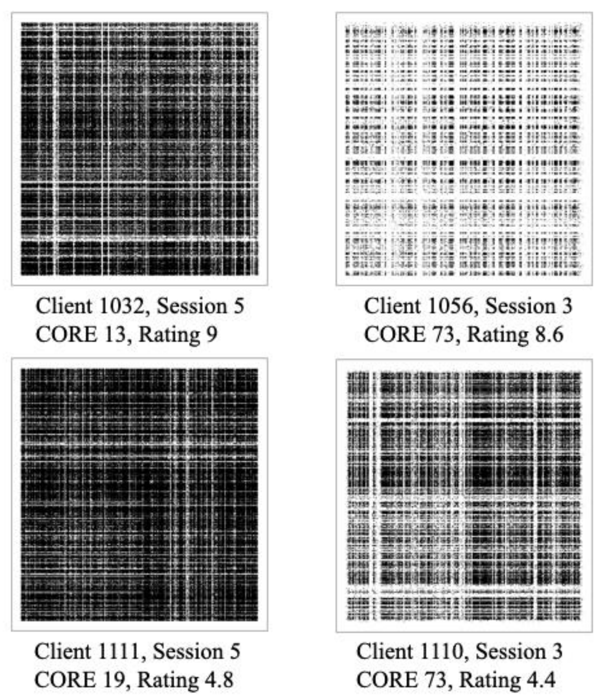

Figure 1: Recurrence plots for session with low client distress and high session quality rating (top left panel), high client distress and high session quality rating (top right panel), low client distress and low session quality rating (bottom left panel) and high client distress and low session quality rating (bottom right panel).

To confirm the link between mental health state and amount of talk, we fitted a multi-level regression with either distress (CORE) or depression (PHQ) score as predictor and overall speech sound duration as dependent variable. As all multi-level models with uncorrelated random slopes of CORE or PHQ scores resulted in singular fit we fitted intercept-only models which confirmed a positive link between PHQ score and speech sound ($\beta = -107.70$, $t = -2.45$, $p = .02$) but not between CORE and speech sound ($\beta = -93.59$, $t = 1.83$, $p = .07$). This shows that clients' severity of depression, but not general distress, was linked to less talk in the counseling sessions.

To ascertain how to best predict session outcome quality ratings from recurrence metrics while controlling for client mental health state and amount of talk we checked for Pearson correlations between recurrence metrics, treating sessions as independent events. The correlation coefficients ranged from .50 to .99 ($n = 239$) and were all significant after Bonferroni correction. To explore the unique contribution of qualitatively different aspects of recurrence we fitted multiple multi-level regression models with each recurrence metric as a separate predictor.

Recurrence metrics were entered in addition to one of the measures of client mental health state and total speech sound duration per session. All models with random slopes either failed to converge or resulted in singular fit leading us to adopt the intercept-only model *lmer (Session Rating ~ Mental Health Score + Sound in Session + Recurrence Metric + (1 | Therapist_ID / Client_ID)*. Table 2 shows coefficients for models fitted to session quality ratings of the therapists (Panels A and B) and the clients (Panels C and D) controlling for either overall distress (Panels A and C), or depression (Panels B and D). We also performed a second set of analyses on the shuffled timeseries to control for incidental recurrence artefacts, which confirmed the predictive effects of the CORE and the PHQ as well as of the total sound duration but showed no significant effects of any of the recurrence metrics (all data and code for main and shuffled analyses available at https://osf.io/n9wcx/files/).

Table 2: Model coefficients for predicting ratings of session outcome quality with centered fixed effects of client mental health state, sound duration and target RQA metric.
* p < 05, ** p < .01, *** p < .001

Panel A: Therapist Rating, Distress

Model	1	2	3	4	5
intercept	5.33*	5.34***	5.34***	5.34***	5.33***
CORE	-0.27***	-0.27***	-0.27***	-0.26***	-0.26***
sound	0.13*	0.13*	0.13*	0.11	0.11
DET	-0.05				
L		-0.05			
ENTR			-0.06		
LAM				-0.13*	
TT					-0.13*

Panel B: Therapist Rating, Depression

Model	6	7	8	9	10
intercept	5.34***	5.34***	5.34***	5.34***	5.34***
PHQ	-0.15*	-0.16*	-0.16*	-0.14*	-0.14*
talk dur	0.13*	0.13*	0.13*	0.11	0.11
DET	-0.06				
L		-0.06			
ENTR			-0.07		
LAM				-0.13*	
TT					-0.12

Panel C: Client Rating, Distress

Model	11	12	13	14	15
intercept	6.73***	6.73***	6.73***	6.74***	6.73***
CORE	-0.28**	-0.29**	-0.29**	-0.29**	-0.30**
sound	0.04	0.03	0.03	0.02	0.03
DET	-0.16*				
L		-0.22*			
ENTR			-0.26**		
LAM				-0.19*	
TT					-0.19*

Panel D: Client Rating, Depression

Model	16	17	18	19	20
intercept	6.75***	6.75***	6.75***	6.76***	6.75***
PHQ	-0.15	-0.15	-0.15	-0.14	-0.15
sound	0.05	0.04	0.04	0.03	0.04
DET	-0.14*				
L		-0.17*			
ENTR			-0.18*		
LAM				-0.18*	
TT					-0.14

Discussion

Our results showed that ratings of session outcome quality were negatively affected by client distress: The higher the CORE score the lower did therapists and clients rate the outcome of the session. Interestingly, PHQ scores, which indicate severity of depressive symptoms and were highly correlated with CORE scores (Spearman's $r = .74$, $n = 225$, $p < .001$), only affected therapists', but not clients' ratings. It appears that a broad indicator of distress across a range of mental health problems either affected clients' communicative behavior to a greater extent or biased their outcome ratings more. While this is an interesting finding to pursue further it was not the primary focus of this study and is only of relevance insofar as it shows that clients' mental health state affected how sessions were evaluated.

Furthermore, the results showed that therapists', but not clients' ratings of session quality could also be predicted from how much talk had taken place. As our analyses did not differentiate between therapists' and clients' speech this association may reflect either that therapists place greater value on conversational contributions of the clients or on greater interaction between the clients and themselves.

While amount of talk seemed to be an important dimension that influenced therapists' evaluation of session quality, clients did not rely on this information in their evaluations. Instead, their ratings were predicted by recurrence metrics, which all had a significant negative effect over and above mental health state and sound duration (aside from Trapping Time in the model that included PHQ as mental health state predictor). For the therapists, only laminarity, i.e., to what extent ISIs remained the same over periods of time, was negatively linked to their ratings. Our finding that therapists base their evaluations on aggregate quantitative measures such as amount of talk while clients evaluate interactions more positively and are more sensitive to qualitative features of the interactional dynamics is in line with findings by Reuzel et al. (2013). That study predicted evaluations of interactions between support staff and learning-disabled clients by independent staff and client observers from recurrence in gaze direction and speech rhythm, and found greater sensitivity of staff evaluations to aggregate quantitative measures like amount of talk and greater sensitivity of clients' evaluations to patterns of recurrence, which in this case were positively linked to evaluations. Thus, participants in psychotherapy and in broader support-type interactions appear to evaluate communicative success differently presumably because basic support and therapy interactions have different goals. Still, our findings confirm that clients are more in tune with dynamic patterns of recurrence in non-verbal features of communication than staff or therapists.

If we accept that client-based evaluations carry greater weight due to their stronger link with treatment outcome (it is likely to matter more what the client thinks about session quality) the consistent negative link with recurrence metrics in this study is an important finding: It suggests that in psychotherapy, clients view a greater degree of recurrence as detrimental to counselling success. This finding confirms the finding by Reich et al. (2014) of a negative link between pitch synchrony and ratings of therapeutic alliance. Placed in the broader context of research on how alignment and coordination affect communicative outcomes our findings show that the direction of this link may strongly depend on communicative context and communicative goal: While alignment and coordination of various features of communication may be predictive of positive outcomes in joint action contexts they may be counterproductive in contexts where communication, among other things, aims to modify entrenched patterns of thinking and behavior.

While our findings suggest that RQA may be a promising avenue for automated appraisal of communicative outcomes in the context of psychotherapy it is important to point to several caveats associated with this study that should be remedied in further research.

First, although the size of our corpus was large compared to other studies in this area, the outcome measures we had to work with were not optimally designed to appraise quality of psychotherapy. Future studies should select better instruments for therapists and clients to evaluate session quality based on theoretically meaningful constructs. More importantly, the predictive validity of RQA metrics of non-verbal parameters of communication would be considerably enhanced by independent ratings of session quality from trained observers.

Second, our study was confined to the analysis of speech rhythm but other non-verbal features of communication such as pitch patterns, body and head movements (Ramseyer & Tschacher, 2014) or heart rate fluctuations (Kodama, Tanaka, Shimizu, Hori et al. 2018) should be considered as well to gain a better understanding of the interplay between different levels and modalities of social interaction.

Third, technical constraints and capacity limitations precluded diarization of the audio-recordings and the application of cross-RQA. While there are theoretical reasons to treat the dyad as a synergistic unit and use auto-RQA as we did (Fusaroli & Tylén, 2016), it would still be important to explore patterns of synchronization between interlocutors using cross-RQA. From the analyses conducted here we are unable to determine whether recurrence occurred within or across speakers; hence, we cannot draw conclusions as to whether the recurrence patterns that negatively impacted clients' session appraisal were due to repetitiveness and perseveration in the client's own speech or the mimicking of speech rhythm patterns between client and therapist.

Fourth, we presented exploratory findings from an observational study which needs corroboration in experimental research to carefully manipulate communicative contexts and goals. For example, mood induction can be employed prior to communication to examine how the interplay between affective states of interlocutors and temporal dynamics of non-verbal speech parameters measured through RQA affects appraisal of attainment of various communicative goals.

Finally, the strong correlation between the various recurrence metrics poses problems for RQA of complex time-series data: Entering highly correlated predictors into regression models leads to multicollinearity yet reducing dimensionality through PCA could obscure individual contributions of, and potential interactions between, these metrics, e.g., amount and length of recurring patterns (DET and L) vs. amount and length of instances a dyad remains in the same state (LAM and TT). At the same time, performing multiple analyses using individual recurrence metrics as predictors can inflate the likelihood of spurious effects. One possible solution that we are currently pursuing is to employ machine learning to predict communicative outcomes directly from recurrence plots treated as visual objects – an approach that may better capture subtle differences in the complexity of recurrence patterns while preserving information from aggregate measures such as amount of talk.

Despite these caveats our findings show that application of RQA should be developed further as a promising tool for automated appraisal of quality of communicative interactions in contexts like psychotherapy, where such appraisals are of great practical interest. Using RQA would also improve the theoretical understanding of which specific patterns of alignment and coordination are indicative of attainment of different communicative goals in a diverse range of social interactions.

Acknowledgments

Pilot work for this project (not reported here) was funded by UK Defence and Security Accelerator (DASA) Grant ACC6005035. We also would like to thank all therapists and clients who agreed to share their session recordings as well as Erin McCormick and Leslie Blaha for providing their RQA-helper function code.

References

Abarbanel, H. D. I. (1996). *Analysis of observed chaotic data.* New York: Springer.

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*(1), 59-69.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. *URL http://www. praat. org/. Retrieved March*, *14*, 2018.

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Psychology*, *5,* 510.

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Psychology*, *5,* 510.

Cooper, M. and McLeod, J. (2011) *Pluralistic counselling and psychotherapy.* Sage London

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385-390.

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PloS one*, *12*(6), e0178140.

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., McGrath, G., Connell, J., Audin, K. (2000). Clinical Outcomes in Routine Evaluation: The CORE-OM. *Journal of Mental Health, 9,* 247–255.

Flemotomos, N., Martinez, V. R., Chen, Z., Singla, K., Ardulov, V., Peri, R., ... & Narayanan, S. (2021evlau). am I A good therapist?" automated evaluation of psychotherapy skills using speech and language technologies. *CoRR*, *abs/2102.11265*.

Ellgring, H. (2007). *Non-verbal communication in depression.* Cambridge University Press.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145-171.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931-939.

Kodama, K., Tanaka, S., Shimizu, D., Hori, K., & Matsui, H. (2018). Heart rate synchrony in psychological counseling: a case study. *Psychology*, *9*(07), 1858.

Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001) The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine, 16*(9) 606-613.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

National Institute of Health and Care Excellence (NICE) (2009) Depression. Available at https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance#treatment-choice-based-on-depression-subtypes-and-personal-characteristics [Accessed 31. May 2020]

Paulick, J., Deisenhofer, A.-K., Ramseyer, F., Tschacher, W., Boyle, K., Rubel, J., & Lutz, W. (2018). Nonverbal synchrony: A new approach to better understand psychotherapeutic processes and drop-out. *Journal of Psychotherapy Integration, 28*(3), 367–384.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284.

Ramseyer, F., & Tschacher, W. (2014). Nonverbal synchrony of head-and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in Psychology*, *5*, 979.

Reich, C. M., Berman, J. S., Dale, R., & Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *Journal of Social and Clinical Psychology*, *33*(5), 481-494.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29-46.

Reuzel, E., Embregts, P. J., Bosman, A. M., Cox, R., van Nieuwenhuijzen, M., & Jahoda, A. (2013). Conversational synchronization in naturally occurring settings: A recurrence-based analysis of gaze directions and speech rhythms of staff and clients with intellectual disability. *Journal of Nonverbal Behavior*, *37*(4), 281-305.

Smith, K. & De La Prida, A. (2021) *A Primer for Pluralistic Counselling and Psychotherapy*. PCCS Books

Takens, F., Rand, D. A., & Young, L. S. (1981). Detecting strange attractors in turbulence. In D. A. Rand & L.-S. Young (Eds.), *Dynamical Systems and Turbulence.* Berlin: Springer.

Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. *Frontiers in Psychology, 9,* 2232.

Webber Jr, C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, *94*(2005), 26-94.

Webber, C. L., & Zbilut, J. P. (2005). Recurrence Quantification Analysis of Nonlinear Dynamical Systems.

In M. Riley, & G. Van Orden (Eds.), *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences* (pp. 26-94). USA: National Science Foundation.