Universität Basel

Fakultät für Psychologie

# Assessment of affective and conversational trajectories in psychotherapy with adolescents suffering from borderline personality disorder

**Inauguraldissertation** zur Erlangung der Würde eines Doktors der Philosophie vorgelegt der Fakultät für Psychologie der Universität Basel von

**Lukas Elias Fürer**

aus Sissach (BL), Schweiz

Basel, 2020

Universität Basel

Fakultät für Psychologie

Genehmigt von der Fakultät für Psychologie auf Antrag von

Prof. Dr. Klaus Schmeck

Prof. Dr. Jens Gaab

Datum des Doktoratsexamen: April, 2020

Dekan/in der Fakultät für Psychologie

**Erklärung zur wissenschaftlichen Lauterkeit**

Ich erkläre hiermit, dass die vorliegende Arbeit ohne die Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel selbstständig verfasst habe. Zu Hilfe genommene Quellen sind als solche gekennzeichnet. Die veröffentlichten oder zur Veröffentlichung in Zeitschriften eingereichten Manuskripte wurden in Zusammenarbeit mit den Koautoren erstellt und von keinem der Beteiligten an anderer Stelle publiziert, zur Publikation eingereicht, oder einer anderen Prüfungsbehörde als Qualifikationsarbeit vorgelegt. Es handelt sich dabei um folgende Manuskripte:

- **Fürer, L.**, Schenk, N., Roth, V., Steppan, M., Schmeck, K., & Zimmermann, R. (2020). Supervised speaker diarization using random forests: a tool for psychotherapy process research. *Frontiers in Psychology*.
- Zimmermann, R., **Fürer, L.**, Schenk, N., Koenig, J., Roth, V., Schlüter-Müller, S., ... & Schmeck, K. (2020). Silence in the psychotherapy of adolescents with borderline personality pathology. *Personality Disorders: Theory, Research, and Treatment*.
- Steppan, M., Zimmermann, R., **Fürer, L.,** Southward, M., Koenig, J., Kaess, M., Kleinbub, J. R., Roth, V., Schmeck, K. (submitted). Machine Learning Facial Emotion Recognition in Psychotherapy Research: Evidence supporting a new technology. *Translational Psychiatry*.
- Schenk, N.***, Fürer, L.***, Zimmermann, R., Steppan, M., Schmeck, K. (in press). Alliance Ruptures and Resolutions in Personality Disorders. *Current Psychiatry Reports*. (* equal contribution)

Basel, Dezember 2020

Lukas Fürer

# Lukas E. Fürer


| | | |
|---|---|---|
| Kontakt | Lukas E. Fürer<br>Wattstrasse 8<br>4056 Basel - CH | *Tel:* +41(0) 79 451 72 40<br>*eMail:* lukas.fuerer@posteo.net |


| | | |
|---|---|---|
| Bürgerort | Sissach BL - Schweiz | |
| Geburtsdatum | 17.03.1990 | |
| Zivilstand | verheiratet | |
| Ausbildung | **Institut für Kinder-, Jugendlichen- und Familientherapie**,<br>Ausbildung zum eidg. dipl. Psychotherapeuten | seit 2017 |
| | **Universität Basel**, Fakultät für Psychologie, MSc in Psychologie<br>Vertiefungsrichtung: Klinische Psychologie<br>Masterarbeit in den Bereichen Psychophysiologie und Persönlichkeitsstörung | 2014 - 2017 |
| | **Universität Zürich**, Philosophische Fakultät, BSc in Psychologie<br>Hauptfach: Psychologie<br>Nebenfach: Neuroinformatik | 2010 - 2013 |
| Schulen | Gymnasium Liestal, Bilinguale Matura: Englisch | 2005 - 2008 |
| berufliche Tätigkeiten | **Doktorat & Klinische Anstellung**, Forschungsabteilung der Kinder- und Jugendpsychiatrischen Klinik Basel (KJPK)<br>Forschungsgruppe Prof. Schmeck<br>Schwerpunkt: Affective Computing, Persönlichkeitsstörungen | seit 2017 |
| | **Zivildiensteinsatz**, Tageshaus für Obdachlose<br>Stiftung Sucht | 2017 |
| | **Forschungsassistenz**, Forschungsabteilung der Kinder- und Jugendpsychiatrischen Klinik Basel (KJPK)<br>Forschungsgruppe Prof. Schmeck<br>Schwerpunkt: Persönlichkeitsstörungen im Jugendalter | 2016 - 2017 |
| | **Applied Behavior Analysis (ABA)**<br>Verhaltenstherapeutische Förderung bei Autismus-Spektrum-Störungen<br>wöchentlich, zweimonatlich supervidiert | 2015 - 2017 |

**Klinisches Praktikum**, Universitäts-Kinderspital Zürich 2014
Abteilung für Neuropsychologie

**Zivildiensteinsatz**, Forschungsassistenz 2014
Abteilungen Kognitive und Molekulare Neurowissenschaften
Universität Basel

**Forschungspraktikum**, Schlafforschung 2013
Institut für Pharmakologie und Toxikologie
Universität Zürich

**Data Management**, Corris AG - ZH 2010 - 2013

**Pfleger**, Altersheim Jakobushaus, Thürnen - BL 2009 - 2010

**Pflegehelfer SRK**, Spitalschulen 41, Moudon - VD 2009

## Sprachen

**Deutsch**, Muttersprache

**Englisch**, sehr gute Kenntnisse in Wort und Schrift

**Französisch**, gute Kenntnisse in Wort und Schrift

**LaTeX**, Textverarbeitung

**Matlab & Python**, Berechnungen, Signalverarbeitung

**R**, Statistiken, graphische Darstellung

Grundkenntnisse in **Java**

**Table of Contents**

## Research Gap

Recently, machine learning methodologies and affective computing have become more popular in the field of psychotherapy research (Aafjes-van Doorn et al., 2020; Poria et al., 2017). Methods facilitate video, audio and text processing in order to extract verbal (Goldberg et al., 2020) and nonverbal information (facial expression: Arango et al., 2019; postures: Zhang et al., 2018; paralanguage: Crangle et al., 2019). A promising application for automated streams of nonverbal information is emotion recognition (Halfon et al., 2020; Sharma & Dhall, 2021). The drastic rise of possibilities and the eclectic use of symbolic information streams begs for anchoring theory to inform meaning making of nonverbal phenomena. Most publications study the mutual interdependence of nonverbal signals in patient and therapist (synchrony, attunement, concordance, coordination), taking a firm stand in generalizing theories (common factor theory, communication theory, information processing theory, self-organisation theory), implying that studied processes of nonverbal exchange translate to transtheoretical concepts of psychotherapy research and human interaction in general (Koole & Tschacher, 2016; Laroche et al., 2014; Salvatore et al., 2015). On a more fine-grained conceptual level, empathy, emotion regulation and quality of therapeutic (working) relationship (alliance) are proposed as correlates (Imel et al., 2014; Reich et al., 2014; Soma et al., 2020). In this dissertation, I illuminate the use of nonverbal signals from a different perspective. Being a practicing therapist myself, I anchor my thoughts in a clinical standpoint, applying theory on mechanisms of change in relational psychoanalysis (Adolescent Identity Treatment: Foelsch et al., 2014; Transference Focused Therapy: Levy et al., 2006) relevant to the psychiatric understanding of (borderline) personality disorders. By collapsing assumptions of relational psychoanalysis (Tufekcioglu & Muran, 2014) and the events paradigm (Timulak, 2010), psychotherapy is conceptualized as the mutual communication and regulation of self-states in an ongoing negotiation process that brings forth key events. More precisely, using object-relations theory (Kernberg, 1995), key events can be defined as episodes of integrative work with the dominant object relation dyads. The instrumental involvement of nonverbal communication streams in this process is discussed (Dreyer, 2018; Mac Cormack, 1997). On the conceptual level, it is possible to find common ground between emotion regulation (Campos et al., 2011) and alliance negotiation (Eubanks et al., 2019)

as they both emerge as nonverbal communication acts in moments of therapeutic work, moving towards the clinical goal of identity integration (Jung et al., 2013; Schlüter-Müller et al., 2020; Schmeck et al., 2013). I discuss how the automatic assessment of nonverbals can help recognizing these key moments. Interpersonal coordination processes have been proposed as correlates for emotion regulation and alliance negotiation. I highlight problems with the concrete adaptation of coordination assessment under the assumptions of the events paradigm and how they can be overcome by the perspective of relational approaches.

## Theory

**Relational psychoanalysis** integrates theory that offers a relational perspective on intra- and interpersonal processes governing and describing the self (Mitchell & Aron, 1999). Muran (2002) summarizes basic tenets (*in italic*) of the relational perspective concerning the structure and operations of the self: *Implicit representational content of self - other interactions* (internal objects: Hinshelwood, 1997; relational schemas: Baldwin, 1992; object-relation dyads: Clarkin et al., 2007) are posited as basic building blocks of the self. They are considered to develop on the basis of interactions with significant others in order to increase the likelihood of maintaining nurturing relationships (Beebe et al., 1997). They are active in expressive-motor and emotional memory stores informing expectancies and strategies for relating with others (Safran et al., 1990). Their learned nature allows for constant adaptation through dialogical interaction (Brokerhof et al., 2018; Centonze et al., 2020). These representational structures are protected by the *interplay between the structures and the processes of the self*. Depending on the therapeutic school, this interplay symptomatically unfolds in cognitive distortions (Geiger et al., 2014), behavioural patterns (Kaess et al., 2013) or defence mechanisms (Zanarini et al., 2009). The representational content gives rise to diverse *self-states* that can be experienced in both conscious (or preconscious) thoughts, feelings and images (Dreyer, 2018), or through enactments of unconscious (transferential) material (Lyons-Ruth, 1999). Transitions between self-states vary on a continuum of seamlessness (Golynkina & Ryle, 1999; Ego States Therapy: Watkins & Watkins, 1997). In dyadic encounters, *self-states are communicated between interlocutors and mutually regulated* (Beebe & Lachmann, 1998). Interpersonal encounters, including dyadic psychotherapy, are thereby characterized

by an ongoing negotiation involving the «push and pull of respective self-states of patient and therapist […] involving the continuous press of both patient and therapist needs for agency and relatedness» (Muran, 2002). The idiosyncratic expression of self-states serves the function of informing the environment about the person and vice versa (Muran & Safran, 2002). This ongoing negotiation is studied in the framework of **psychotherapy process research** and the **events paradigm** (Greenberg, 1991; Rice & Greenberg, 1984). Process research untangles how psychotherapy works by focusing on in-session variables such as therapist or patient behaviour and their interaction during treatment (Castonguay & Beutler, 2006). The events paradigm defines delineated therapeutic events linked to mechanisms of change (Silberschatz, 2017). They are translated into observable behaviour, whose occurrence, frequency and timing are linked to outcome (Buchholz, 2019; Elliott, 2010). Concentrating on key moments, rather than session level averages, these moments hold the pivotal benefit of being describable to and identifiable by therapists, potentially fostering training and process monitoring.

As discussed, the relational lens considers psychotherapy as an ongoing process of mutual communication of self-states and regulation acts thereof (Centonze et al., 2020). **Nonverbal behaviour** is instrumental in the communication and mutual regulation of self-states (embodiment theory: Tschacher & Storch, 2012). Just as in a play (theatre metaphor: Hermans, 2006; Mac Cormack, 1997), states are assumed by modulation of nonverbals. For example, Dreyer (2018) showed how the voice is modulated in play therapy according to the roles assumed. The modulation of the nonverbal serves two functions: 1) It allows to reactivate the self-state, to fully experience it, and 2) to communicate the intensity and quality of the self-state to the interlocutor (Armstrong et al., 2015). In order to move forward in the story, in order for the interaction to be ongoing and mutual, the state has to be checked out by the interlocutor, modulating its nonverbals in response to the partners state changes. Thereby changes in the nonverbals of the interlocutor again serve to communicate that the state switch has been recognized and acknowledged (Grace et al., 1995). This process unfolds both consciously and unconsciously in transferential material (Böhmer, 2010; Normandin et al., 2015) and across different signals and time scales. Literature on the **Interpersonal Coordination** (IC) of nonverbal behaviour assumes, that a smooth and coordinated communication, recognition and acknowledging of self-states, as reflected in nonverbal signals, brings forth the experience of being understood (empathy: Imel et al.,

2014), of being on the same wavelength (relationship quality: Ramseyer & Tschacher, 2011), and of being held (emotion regulation: Soma et al., 2020; attachment: Håvås et al., 2014).

In psychiatric diagnostic terminology, the life-long adaptive processes of self- and other relatedness are caught under the umbrella term of self- and interpersonal functioning and, altogether, as **personality functioning** (Birkhölzer et al., 2015, 2020; Goth et al., 2018). **Borderline Personality Disorder** (BPD) is a severe mental illness that, at its core, is characterized by deficits in self- and interpersonal functioning. Personality functioning is dimensionally measured in self-functioning dimensions 'identity', 'self-direction' and interpersonal functioning dimensions 'empathy' and 'intimacy' (Goth et al., 2018). This dissertation evaluates data gathered in the therapeutic context of **Adolescent Identity Treatment** (AIT; Foelsch et al., 2014). As the name suggests, Adolescent Identity Treatment posits deficits in identity functioning to be at the core of the broad maladaptive behaviour, deficits in self- and interpersonal functioning and negative affective experience in borderline pathology (Schlüter-Müller et al., 2015). The development of a stable and coherent identity is a crucial developmental task in adolescence (Schmeck et al., 2013; Sollberger, 2013). **Identity disturbance (identity diffusion)** – in contrast to non-pathological identity crises – is a psychiatric syndrome marked by a subjective sense of lack of coherence (Foelsch et al., 2010; Schlüter-Müller et al., 2015; Wilkinson-Ryan & Drew, 2000). It emerges as a key feature in the diagnosis of personality pathology (Benzi & Madeddu, 2017; Foelsch et al., 2010; Jung et al., 2013; Kernberg, 2006; Sollberger et al., 2012, 2015). Identity disturbance has been associated with the presence and severity of (borderline) personality pathology in adults and adolescents (Feenstra et al., 2014; Jung et al., 2013; Lind et al., 2019; Sollberger et al., 2012) and it has been shown to fully mediate the effect of mentalization deficits on interpersonal problems (De Meulemeester et al., 2017). Further, using exploratory bi-factor analysis on the categorical items, Sharp et al. (2015) showed that the item 'identity disturbance' correlated highest with a 'g'-factor of personality disorder.

Kernberg's **Object Relations Theory** is one specific theory of relational psychoanalysis (Kernberg, 1995). In object relations theory, single units of representations of self and other, linked by an affective state, are called object relations dyads (Levy et al., 2006). The activation of object relations dyads is accompanied by a shift in self-state. Identity disturbance is characterized by a pathological

organisation of polarized and non-integrated object relations dyads (Gagnon et al., 2016). Patients engage dissociative and splitting defences and reinforcement strategies (projective identification, devaluation, denial, idealization) to protect idealized from malignant object relations (Kernberg et al., 2008). Internal representations of the self and significant others consequently are separated into positive and negative aspects that are actively held apart, implying that only one can consciously be held at a time (Kernberg, 2006). Clinically, patients with identity diffusion present with the inability to describe the self and others in detail and they struggle with the establishment of close, meaningful and intimate relationships (Levy et al., 2006). Identity diffusion was a main inclusion criterion for the study sample analysed in this dissertation. Adapted from **Transference Focused Therapy** (Yeomans et al., 2013), AIT uses **clarifications, confrontations and interpretations of transference and countertransference reactions** in the here and now in order to work on the dominant and activated object relations in an effort to bridge opposite representations of the self and others (Foelsch et al., 2014). Psychotherapy thereby is formulated as an act of identity integration. As is evident from the description of object relations, this integration process is accompanied by affective expression and can cause anxiety because it involves moving towards previously denied aspects of self (Yeomans et al., 2013). Therapists are invited to provide a holding environment that allows patients to reactivate dangerous object relations dyads and experience related (mostly negative & aggressive) affects without the overwhelming quality of the past, knowing that therapists can tolerate the negative affective states (Levy et al., 2006). The identification and exploration of affective states further serves the aim to shed light on the dynamics of extra- and the intratherapeutic relationship (Yeomans et al., 2013). The integration process can cause anxiety and patients employ defences in order to prevent the activation of object relations dyads and in turn keep the related affects at bay (Clarkin et al., 2007). Although the therapeutic relationship is previously protected by controlled conditions (treatment contract), these defences may be directed towards the relationship (Safran & Muran, 1996). Concerning affect activation, Greenberg & Safran (1989) aptly differentiate between primary and secondary emotions. Secondary emotions can be results of a responsive and defensive self and are directed towards the other (therapist), implying that the responsibility lies with the other. They can be understood as reactions to the primary adaptive emotion

surfacing through the activation of the object relations dyad. Primary emotions in turn are characterized by acceptance and responsibility for one's feelings.

Both the activation of (peak) affective states and the occurrence of interpersonal defences, threatening the therapeutic relationship, can thereby be found to intermingle as they result from a defensive self during working process of identity integration. In accordance with core difficulties in self and interpersonal functioning and related increased maladaptive defences (Bond et al., 1994), it has been hypothesized that psychotherapy with patients with borderline personality disorders is, to a larger extent as in other patients, characterized by momentary strains in the therapeutic relationship, and brings about increased demands in mutual affect regulation efforts during moments of affect activation.

As described above, this process of identity integration – potentially threatening the therapeutic relationship and involving affect activation – unfolds in an ongoing negotiation process over time. Both affect activation and processing and negotiation of the therapeutic relationship have been intensively studied in form of proposed key moments. I will introduce relational definitions of both emotion regulation and relationship negotiation.

Evidence amounts that the exploration of emotions is found across therapeutic schools and is related to good outcome therapies (Greenberg & Paivio, 2003; Greenberg & Pascual-Leone, 2006; Peluso & Freund, 2018). **Emotional dysregulation** is of interest across disorders and is found to be a major vulnerability for mental disorders (Kring, 2008). Difficulties in emotion regulation are considered a core concept in the development and maintenance of psychiatric disorders and emotion regulation is discussed as a transdiagnostic treatment construct (Sloan et al., 2017). A multitude of scales and rating systems approach affect activation and processing in the psychotherapeutic encounter (Client Emotional Arousal Scale: Carryer & Greenberg, 2010, Client Experiencing Scale: Pascual-Leone & Yeryomenko, 2017), Client Emotional Productivity Scale: Auszra et al., 2013). **Emotion Regulation** is considered a cornerstone mechanism of psychotherapy. In contrast to intrapersonal regulation based on demarcated regulation strategies (D'Agostino et al., 2017), relational approaches see the regulation process as an ongoing mutual regulation process with others (Sloan et al., 2017). But what exactly is regulated and to what end? Campos et al. (2011) argue that it is not an emotional state per se that is regulated, but rather "a conflict between the goals of one person and those of another, and, on occasion, a conflict between

the goals of a single person". The regulatory mechanism is coined as a goal negotiation that moves toward the end of a negotiated outcome. This outcome does not need to involve feeling better, ergo being effective, but needs to be adaptive, including a coordination of future concerns. This definition presents emotion regulation feasible to be studied as a momentary interactional phenomenon that evolves around and through the mutual communication of self-states. The emotion per se takes a back seat being thought of as the by-product of a negotiation process striving for a persistent fit of individual needs and goals of both patient and therapist.

Another transtheoretical concept that has been studied through the relational and key moments lens is the **Alliance**. The alliance indicates the "quality and strength of the collaborative relationship between client and therapist" (Norcross & Hill, 2002). As a common factor, alliance is theorized to be jointly responsible for comparable effectiveness of different treatments with diverse theoretical backgrounds and derived techniques (Laska et al., 2014; Rosenzweig, 1936; Wampold & Imel, 2015). A vast array of instruments has been presented to measure alliance, Horvath (2011) estimates over 70, stemming from different theoretical backgrounds and holding all imaginable forms. The relational approach has coined alliance as an ongoing intersubjective negotiation process that aims at resolving momentary deteriorations in the alliance (Muran, 2002; Safran, 1993; Safran & Muran, 1996). Non-collaborative patient behaviour regarding therapy goals and tasks or strains in the affective bond, termed ruptures, are sought to be repaired in resolution efforts (Lingiardi & Colli, 2015). The Rupture and Resolution Rating system (3RS; Eubanks et al., 2019) offers the concurrent operationalization. Two types of ruptures are defined: withdrawal and confrontation rupture. Withdrawal ruptures are characterized by a patient's movement *away* from (minimal response, avoidant storytelling, self-criticism/hopelessness) or *towards* (appeasing) the therapist and/or the work of therapy. Confrontation ruptures involve the patient moving *against* the therapist and/or work of therapy (complaints/concerns about the therapist or progress in therapy, rejecting therapist intervention, efforts to control/pressure therapist). Resolution strategies can be either 'immediate' (providing a rationale for treatment, clarifying misunderstandings, changing tasks or goals or providing validation for defensive behaviour) or 'expressive', focusing on exploring the patients' core relational themes (Muran & Safran, 2017).

It has become evident under relational tenets, that both emotion regulation and alliance negotiation are acts of reacting in an ongoing communication of self-states, signalling the recognition of state changes (i.e., activations of object relation dyads). This back and forth is partly conveyed in nonverbals. Therefore, the study on the quality and extent of interdependence of nonverbals has gained much attention and provides a reasonable avenue for the study of beforementioned mechanisms. However, I will outline some problems that arise when thinking about interpersonal coordination under the event paradigm.

Human interaction in general is characterized by behaviour matching and **Interpersonal Coordination** (IC), IC being the "smooth meshing of interaction" over time (Bernieri & Rosenthal, 1991). Processes of IC are the swiss army knife of psychotherapy research and related fields. Having been identified as correlate of a multitude of relevant concepts, they have moved to the centre of psychotherapy research because of their integrative power (Dales & Jerry, 2008; Koole & Tschacher, 2016). However, to the same extent that they are handy, they can equally be elusive. They present in a multitude of verbal and nonverbal signals and on a multitude of time scales. As described above, IC literature assumes, that a smooth and coordinated communication and recognition of self-state related nonverbals, where interlocutors are responsive towards each other, is instrumental to the experience of empathy, emotion regulation and to the quality of the relationship. With coordination being hypothesized to be a central concept across many theoretical concepts, no validity checks have yet been provided. For example, when interested in the correlation of coordination and emotion regulation, it is imperative to pick events of affect activation and compare them to coordination in moments of nonactivation of affects, in order to definitely link coordination in a signal to the proposed concept. In addition, publications often fail to deliver convincing underlying mechanisms. As I have proposed in the sections before, making clear cut distinctions between theoretical concepts (emotion regulation or alliance negotiation) might be confusing. It would be more relatable when considering them overlapping being concrete communication acts in an ongoing process of self-state sharing. Also, the literature ignores the key events structure of therapy (Timulak, 2010). Rather, until now, studies have used session level averaging in order to predict questionnaires on overall relationship and outcome (Paulick et al., 2018; Reich et al., 2014). This ignores the idiosyncratic moment-to-moment dyadic interaction that is

characteristic for psychotherapy (Martinez et al., 2012; Prince, 1997) and prevents clear clinical conclusions. Further, the study of different concepts relies on different kinds of IC (in-phase – synchrony versus out-of-phase – asynchrony) depending on the concept studied. These distinctions can be confusing and contradicting and are not anchored in psychotherapeutic theory. For example, Reich et al. (2014) expected vocal synchrony to positively correlate with the quality of the therapeutic relationship. Surprised by a negative correlation, they struggle to explain results. Again, collapsing both in-phase and out-of-phase IC as being part of the same communicative process allows to dismiss this distinction, rather concentrating on temporal interdependence. Finally, but in my opinion most importantly, studies provide results in separated modalities. However, when assuming that coordination involves the signalling of a recognition of self-state changes, this can also be achieved in a different modality than the state was communicated in. For example, the therapeutic conversation leads to the activation of an object relations dyad that is characterized by deep sadness. The patient expresses this sadness through vocal behaviour. But, in order to react to the state shift, the therapist does not modulate the voice, rather speaking with the same loudness and pitch, yet moves forward with the upper body, changing posture. This also relates to the problem of interpersonal differences in the use of nonverbals across individuals – some speaking with their bodies, others with their voice – which provides us with well appreciated diversity in real life relationships but is troubling in scientific endeavours.

**Conclusions**

I will provide conclusions that are important to structure the here presented and future research efforts:

1) The psychotherapeutic dyad is engaged in the mutual verbal and nonverbal communication and regulation of self-states that is characterized as an ongoing negotiation.
2) Techniques of AIT (clarification, confrontation, interpretation) used in the psychotherapy with BPD patients foster the reactivation of denied (split-off) and potentially dangerous or overwhelming self-states (object relation dyads) in an effort to promote identity integration.

3) Reactivation of denied self-states is inevitably accompanied by activation of affects (primary or secondary) and defensive mechanisms, including defences directed at the therapeutic (working) relationship (alliance).
4) This activation is momentary and presents in key events over the course of therapy (events paradigm).
5) Nonverbal behaviour due to self-state activation allows both to re-experience the self-state and to communicate the intensity and quality of said state.
6) Both emotion regulation and alliance negotiation can be painted in light of being concrete communication acts of self-state sharing, also embodied in nonverbal behaviour.
7) The ongoing process of nonverbal communication and acknowledgement of self-states can be studied by means of interpersonal coordination.
8) Recognising nonverbal behaviour as a communication act in the ongoing negotiation of self-states broadens the scientific perspective on the study of IC: 1) The distinction between in-phase and out-of-phase coordination is dismissed, 2) Key moments, rather than session averages, are proposed to be of interest, 3) Multimodality and interpersonal preferences in nonverbal channels are considered.

## Methods

The research presented here is done under the umbrella of a clinical study that tested non-inferiority of Adolescence Identity Treatment (AIT) against Dialectical Behavior Therapy for adolescents (DBT-A). Data on the outcome and process was collected in two study centers (DBT-A: Heidelberg, Germany; AIT: Basel, Switzerland), but only data from the AIT study branch is analysed due to technical problems in the German centre. In total 23 patients have been included, 6 dropped out early. The following inclusion criteria were applied for the patients: age 13–19 years; three or more BPD criteria (Structured Clinical Interview for DSM-IV Axis II Personality Disorders; First et al., 1997); and identity diffusion according to the Assessment of Identity Development in Adolescence (AIDA; total t score > 60; Goth et al., 2012). Recruitment was open to male and female patients, only two male patients were enrolled. Patients received 25 psychotherapy sessions. All sessions were audio-visually recorded.

## Research Papers

### S1: Supervised Speaker Diarization Using Random Forests: A Tool for Psychotherapy Process Research

The nonverbal communication of self-states can be measured in a multitude of response systems: Nonverbal movement (Dosamantes-Alperson, 1979), posture (Charny, 1966), vocal and conversational parameters (Roessler & Lester, 1976), heart rate (Lifshitz & Blair, 1960), electrodermal activity (Toomin & Toomin, 1975), breathing (Eigen, 1977), eye movements (Owens, 1977). We identified the audio signal as a promising and feasible signal to be analysed for several reasons: 1) It can be noninvasively captured in high sampling frequency, 2) Compared to physiological signals and movement it is understudied with only a handful of studies in the psychotherapeutic setting (Bone et al., 2014; Imel et al., 2014; Reich et al., 2014; Soma et al., 2020), 3) Besides vocal features related to emotional speech (F0, intensity), it allows to study conversational parameters, that are likely to be associated with the ongoing negotiation process (silence, rhythm, length of speaking turns). When aiming to evaluate speech features, one encounters a practical problem: First, speakers have to be separated, creating a feature stream indicating who speaks when (speaker diarization: Joshi et al., 2016). Research on vocal features in psychotherapy so far has made use of manually coding speech turns, which is time consuming (Reich et al., 2014; Soma et al., 2020). In order to facilitate future research in this promising channel of communication, this research gap had to be addressed by providing automation of the problem. Fully automated (unsupervised) methods currently exist; however, they are prone to errors (Xiao et al., 2015). We chose the middle path and devised and validated a method for supervised speaker diarization relying on machine learning methodology (Zhang et al., 2019). The method makes use of a small number of manual codings to diarize whole dyadic psychotherapies. We tested the method on an open-source speech corpus and made it publicly available (Fürer, 2020).

frontiers in Psychology

METHODS
published: 28 July 2020
doi: 10.3389/fpsyg.2020.01726

# Supervised Speaker Diarization Using Random Forests: A Tool for Psychotherapy Process Research

*Lukas Fürer[1]*, Nathalie Schenk[1], Volker Roth[2], Martin Steppan[1], Klaus Schmeck[1] and Ronan Zimmermann[1,3]*

[1] Clinic for Children and Adolescents, University Psychiatric Clinic, Basel, Switzerland, [2] Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland, [3] Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Basel, Switzerland

Speaker diarization is the practice of determining who speaks when in audio recordings. Psychotherapy research often relies on labor intensive manual diarization. Unsupervised methods are available but yield higher error rates. We present a method for supervised speaker diarization based on random forests. It can be considered a compromise between commonly used labor-intensive manual coding and fully automated procedures. The method is validated using the EMRAI synthetic speech corpus and is made publicly available. It yields low diarization error rates (M: 5.61%, STD: 2.19). Supervised speaker diarization is a promising method for psychotherapy research and similar fields.

**Keywords: supervised speaker diarization, psychotherapy process measure, dyadic audio analysis, EMRAI speech corpus, random forest**

OPEN ACCESS

***Edited by:***
*Giuseppe Sartori,*
*University of Padua, Italy*

***Reviewed by:***
*Cristina Mazza,*
*Sapienza University of Rome, Italy*
*Graziella Orrù,*
*University of Pisa, Italy*

****Correspondence:***
*Lukas Fürer*
*lukas.fuerer@posteo.net*

***Specialty section:***
*This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology*

***Received:*** *10 February 2020*
***Accepted:*** *23 June 2020*
***Published:*** *28 July 2020*

***Citation:***
*Fürer L, Schenk N, Roth V, Steppan M, Schmeck K and Zimmermann R (2020) Supervised Speaker Diarization Using Random Forests: A Tool for Psychotherapy Process Research. Front. Psychol. 11:1726. doi: 10.3389/fpsyg.2020.01726*

## INTRODUCTION

Human interaction is organized by interpersonal coordination that manifests itself in temporally coordinated behavior. Interpersonal coordination can be broadly grouped into behavior matching and interpersonal synchrony, which involve the rhythmic and "smooth meshing of interaction" over time (Bernieri and Rosenthal, 1991). During the dyadic interactions of psychotherapy, patients and therapists have been shown to synchronize in verbal, non-verbal, and physiological behavior (Marci et al., 2007; Ramseyer and Tschacher, 2008; Lord et al., 2015; Koole and Tschacher, 2016; Kleinbub, 2017). A growing body of empirical research has associated the degree to which interpersonal synchrony is present during therapy with therapeutic outcome (Ramseyer and Tschacher, 2014), empathy (Marci et al., 2007; Imel et al., 2014; Lord et al., 2015), the formation of the therapeutic relationship (Ramseyer and Tschacher, 2011), personality traits (Tschacher et al., 2018), and emotion regulation (Galbusera et al., 2019; Soma et al., 2019). Due to their integrative value, processes of interpersonal synchrony have thus moved to the center of attention of psychotherapy research and related fields (Ramseyer and Tschacher, 2006). In the case of non-verbal movement synchrony, motion energy analysis has become a widespread tool to quantify movement from video (Ramseyer, 2013). It is made available through standalone software (Ramseyer, 2019), a MATLAB implementation (Altmann, 2013), and an R-package for synchronization analysis and visualization (Kleinbub and Ramseyer, 2019). This allows researchers to engage non-verbal synchrony in an automized, objective, reproducible, and non-labor-intensive fashion in their respective setting and has accelerated research on non-verbal movement synchrony

in the clinical dyad (Delaherche et al., 2012). In the same line, autonomic measures (heart rate, skin conductance, breathing) applied in the field of interpersonal physiology (Kleinbub, 2017) also benefit from accessible measurement in the naturalistic setting (Weippert et al., 2010; Pijeira-Díaz et al., 2016; Barrios et al., 2019). In contrast, studies on vocal quality or vocal coordination have not gained the same amount of attention (Imel et al., 2014; Reich et al., 2014; Tomicic et al., 2017; Soma et al., 2019; Zimmermann et al., 2020). This is somewhat surprising because audio recordings are a widely used tool for educational, scientific, and supervisory activities (Aveline, 1992) and, in comparison to video or physiological measures, are non-invasive and inexpensive to attain in high quality. However, while the processing of non-verbal movement or physiological measures is facilitated through software solutions and devices, post-processing of audio for quantitative statistics can be strenuous due to speaker diarization (Anguera et al., 2012).

## SPEAKER DIARIZATION IN PSYCHOTHERAPY RESEARCH

Speaker diarization is the practice of determining who speaks when (Anguera et al., 2012). In other words, diarization means creating a feature stream indicating speaker identity over time. Diarization in psychotherapy research is currently practiced in two different ways. On one side researchers rely on manual annotation of speaker identity, being time intensive but accurate (Imel et al., 2014; Reich et al., 2014; Soma et al., 2019). On the other side researchers rely on unsupervised automated methods, presenting with a minor work intensity but also with higher error rates (Xiao et al., 2015; Nasir et al., 2017a,b). The term "unsupervised" indicates that the system is not given prior knowledge as to how the speakers are embodied in the audio features. Mostly, the audio stream is segmented into speaker homogenous segments, which then are clustered (Tranter and Reynolds, 2006). In the field of psychotherapy research, studies have used unsupervised methods producing diarization error rates above 10%. For example, Xiao et al. (2015) used automatic speech recognition in motivational interviewing to produce text-based empathy scores of sessions and compare them with human empathy ratings. They employed a clustering based unsupervised diarization procedure that produced an error of 18.1%. Nasir et al. (2017b) predicted the outcome of couple therapy using speech features. The audio stream was segmented to indicate speaker changes based on generalized likelihood ratio criteria, which then are clustered to provide speaker-homogenous segments. Average pitch information in these segments are then used to provide a speaker annotation (wife or husband). They report a diarization error rate of 27.6%. While fully automated diarization procedures are appealing, diarization error rates can substantially be improved when introducing a learning step into the procedure, based on a small quantity of pure data (Sinclair and King, 2013). This relates to the idea of supervised machine learning. A recent study on a new fully supervised speaker diarization method using recurrent neural networks reported an error rate of 7.6% on a corpus of telephone calls (Zhang et al., 2019).

As described, regarding diarization practices in psychotherapy research, researchers tend to rely either on manual coding, which makes research very cost intensive, or they resort to fully automized unsupervised methods. In order to overcome this obstacle and to accelerate scientific undertakings on audio recordings in psychotherapeutic settings, we introduce a method for supervised speaker diarization, developed to work for standard single microphone audio recordings of dyadic talk psychotherapies. Considering the workload, the supervised method is a compromise between work intensive manual annotation and error prone unsupervised methods. It involves creating a learning set and introducing a learning step prior to automatically diarizing the whole data set.

## AIM OF THIS STUDY

The aim of this study is to present a supervised method for dyadic speaker diarization based on a random forest algorithm. The method is tested using a freely available speech corpus. In the future, this will allow testing alternative methods and refinements of the current method on the same data set. The code has been made publicly available (Fürer, 2020). The procedure has been aggregated to one function and the preparations to run the function have been documented. We hope that this allows researchers with minimal coding experience or unfamiliar with MATLAB to carry out analyses on their own. The method is conceptualized in MATLAB and relies on readily available components (Segbroeck et al., 2013; Giannakopoulos and Pikrakis, 2014). We hypothesize that the method will produce diarization error rates comparable to current supervised diarization methods employed in other fields (below 10% per dyad; Zhang et al., 2019). Based on using random forest algorithm, we further hypothesize that the dyadic out-of-bag error rate (explained below) will positively correlate with the dyadic diarization error calculated on a test set. In future studies, this would allow quality checks on a dyadic level without producing a separate test set.

## METHODS

### Random Forest

The presented method for supervised speaker diarization in dyadic psychotherapy is based on a random forest algorithm. While machine learning methods in general have gained attention in psychological research (Orrù et al., 2020), random forests can be considered a rather understandable machine learning algorithm that has already found its way into psychotherapy research (Imel et al., 2015; Masías et al., 2015; Husain et al., 2016; Sun et al., 2017; Wallert et al., 2018; Zilcha-Mano, 2019; Rubel et al., 2020; Zimmermann et al., 2020). The random forest algorithm is a machine learning classifier based on decision trees (Kotsiantis, 2013). The random forest combines a certain amount of decision trees in a single prediction model and is consequently also called an ensemble learner. It can be employed for regression or classification problems. When

confronted with classification problems, the decision is a majority vote over all trees in the ensemble, which, in ensemble format, provides greater accuracy (Breiman, 2001). Major advantages of the random forest algorithm are that it is insensitive to multicollinearity in the input data and to variables that do not contribute to the classification strength (Imel et al., 2015). In our setting this is of importance since we don't know which variables will be important for which dyad, and it is assumed that speech features may be highly correlated. The "random" in random forest refers to the usage of a random subsample of variables and a random subsample of data entries in the learning set when growing each tree (Husain et al., 2016). The process of randomly selecting a subsample of data entries without replacement for the training of each tree is called bagging (Breiman, 2001). This bagging process allows for the calculation of an out-of-bag error rate, which can be considered an estimate for the generalization error (Breiman, 1996). For each entry in the learning set the trees not using this specific entry for learning can be identified. They are called the out-of-bag classifier. The out-of-bag error is the error produced by the out-of-bag classifier, estimated using only the learning set. Given our use case, the possibility to estimate the generalization error with only the learning set is useful: If we apply this method to new and real psychotherapy audio and calculate the out-of-bag error on the learning data, we can estimate the overall strength of the prediction in each dyad, informing us for which dyads the diarization worked well and for which it didn't. We therefore report the correlation of the dyadic out-of-bag error with the dyadic speaker error (explained below) calculated in the separate test set.

## Supervised Diarization in Dyadic Psychotherapy

The dyadic nature of talk therapy allows for an assumption to simplify the otherwise more complicated diarization process: the number of speakers is known, two in this case. Relating to the idea of supervised learning, here, a classifier is given prior knowledge as to how the two speakers are embodied in the input features (supervised diarization). Fortunately, inside the context of psychotherapy research, the classifiers do not have to be generalizable to different dyads, but rather, multiple classifiers can be trained, each one specialized to diarize one dyad only. The necessary steps involve: (1) creation of a learning set for each dyad (human coder), (2) automatic silence detection, (3) automatic voice activity detection, (4) feature extraction, (5) learning to provide a dyadic classifier, (6) prediction in one dyad, and (7) data aggregation. The steps are explained below.

## EMRAI Synthetic Diarization Corpus

The supervised diarization method is tested on the EMRAI Synthetic Diarization Corpus (Edwards et al., 2018). This corpus is based on the LibriSpeech Corpus (Panayotov et al., 2015), namely recordings of English audiobooks. The manual labeling of audio data for training purposes is extremely time intensive. Thus, the authors of the corpus have synthetically created both 2-person and 3-person "dialogues" with and without overlap by sequentially arranging spoken parts. The EMRAI synthetic diarization corpus thereby offers an opportunity for testing diarization systems built for the context of the dyadic conversations as given in talk therapy.

## Silence Detection and Voice Activity Detection

For silence detection, an algorithm calculates an individual intensity threshold value for each session recording. For more information, please refer to the source code (Fürer, 2020). The result of silence detection is a vector indicating silence and non-silence windows in the audio file. In a second step, voice activity detection is performed using a robust and competitive voice activity detection system for MATLAB developed by Segbroeck et al. (2013). This differentiates between voice and noise in the non-silence windows. Voice activity detection was performed over the whole audio, not only in non-silence windows. The procedure feeds contextually expanded spectral cues related to speech (spectral shape, spectro-temporal modulations, harmonicity, and the spectral variability) to a standard Multilayer Perceptron classifier (Segbroeck et al., 2013).

## Feature Extraction

In order to allow the classifier to accurately differentiate between patient and therapist speech, appropriate features need to be extracted from the audio file. We aimed at using an existing and open source MATLAB library to make the procedure replicable by others. Features are provided by the MATLAB Audio Analysis Library and its function "stFeatureExtraction" (Theodoros and Aggelos, 2014). The function yields a total of 35 audio features: energy, zero-crossing rate, entropy of energy, two spectral centroids, spectral entropy, spectral flux, spectral flux roll-off, 13 Mel-frequency coefficients, 12 chroma vectors, harmonic ratio, and mean fundamental frequency. All audio features and their calculations are described in detail in the introductory publication accompanying the library (Giannakopoulos and Pikrakis, 2014). Here, we will focus our description on the Mel-frequency coefficients (MFCCs), since they are crucial features for speaker diarization (Friedland et al., 2009). The calculation of MFCCs takes into account that our perception of the frequency spectrum is not linear (Goldstein, 2010). We perceive differences in lower frequencies as more predominant than differences in higher frequencies. This non-linear relationship is represented by the mel scale, a function which, informed by psychoacoustics, mimics the human auditory system (Zhou et al., 2011). First, the audio signal is represented in the frequency domain by calculating the log discrete Fourier transform. The power spectrum then is submitted to a mel-scale filter bank consisting of overlapping triangular bandpass filters. Their bandwidth and spacing are given by a linear mel scale interval (Umesh et al., 1999). That way, the frequency spectrum is filtered (warped) in the same way, as it is thought to be filtered in the auditory system. MFCCs are then provided as the discrete cosine transform of the mel-filtered log power spectrum, providing coefficients in the time scale (Kathania et al., 2019). The authors of the MATLAB Audio Analysis Library have calculated MFCCs according to Slaney (1998).

In addition to the features provided by the MATLAB Audio Analysis Library, we calculated HF500, being a voice quality ratio between high spectral energy (above 500 Hz up to 3500 Hz) and low spectral energy (80–500 Hz). It has extensively been used in arousal quantification from speech (Bone et al., 2014; Chen et al., 2016).

All features, including silence and voice activity detection, are calculated in non-overlapping windows of 0.1 s in width, and all features have been used in training and predicting the diarization models.

## Learning and Classification

The described features are then used to train a random forest classifier per dyad to predict speaker identity based of the features using the available speaker annotations from the learning set. Only spoken parts (labeled with person-1 or person-2 speech, no silence) were introduced to the learning set. To illustrate, the learning set would contain the timestamps (start of utterance and stop of utterance) and a variable of speaker identity of all utterances for person-1 and person-2 in the first 10 min of the recording. Using the corresponding features, an ensemble of 500 trees is trained for each dyad, using Breiman's algorithm (Breiman, 2001). All EMRAI dialogues of length bigger than 20 min ($n$ = 107) were selected. The first 10 min of each dialogue were chosen for learning purposes, while minutes 10–20 were used as a test set. This simulates the creation of a learning set in a naturalistic setting. Using 10 min of audio in the learning material means that each speaker is represented by less than 5 min of speech (Mean: 4.17 min, Std: 0.38). After training, the classifier is then used to predict speaker identity in the independent test set (minutes 10–20 of the respective recording), resulting in classifications of either person-1-speech or person-2-speech. Please note that the classifier would also yield a decision for actual silence windows; it was not trained to discriminate between silence, noise, and spoken parts. This requires aggregating information to a final decision.

## Data Aggregation

After classification is acquired, three information streams must be aggregated in order to produce a final diarization vector. Results of silence detection (silence or no silence), voice activity detection (voiced or unvoiced), and random forest based diarization (person-1 or person-2 speech) are combined to a feature stream of 0.1s segments of either non-speech, person-1-speech, or person-2-speech according to the following rules: Windows classified as silence by the silence detection remain unchanged. Non-silence windows, however, are replaced by the information stream of the voice activity detection resulting in a combined stream indicating silence, non-speech/noise, and speech. The windows classified as speech are then replaced by the person-1-speech and person-2-speech labels obtained by the respective classifier. The resulting vector contains the labels "non-speech" (silence or noise), "person-1-speech," and "person-2-speech."

## Error Reporting and Data Set

The performance of a speaker diarization method is assessed via the diarization error rate (Barras et al., 2006), a measure comprised of the sum of the following elements: (1) *speaker error* (SpE, percentage of times the wrong speaker is predicted), (2) *missed speech* (MSp, percentage of times silence is predicted instead of speech), (3) *false alarm speech* (FASp, percentage of times speech is predicted instead of silence), and (4) *overlap error* (percentage of times overlapped speech is not assigned to one of the respective speakers). Given our choice of using 2-person non-overlapping speech, the diarization error rate (DER) is reported as the sum of the first three errors (Reynolds and Torres-Carrasquillo, 2005). SpE, MSp, and FASp are reported as mean values with standard deviations over all dyads, same-sex dyads, and different-sex dyads. The sampling frequencies (fs) of the corpus and our prediction stream were different (fs corpus = 100, fs prediction stream = 10) insofar as 10 windows at a time of the corpus are summarized to match one window of our prediction. Transitional windows, where more than one classification was present in the corpus windows to be summarized (both speech and silence), are excluded from the analyses.

We also hypothesized that the dyadic out-of-bag error would be a useful measure to control for the quality of the diarization (speaker annotation) in any specific dyad. We report the correlation between the dyadic out-of-bag error and the dyadic SpE.

# RESULTS

## Total DER, Speaker Error, Missed Speech, False-Alarm Speech

**Table 1** provides an overview over error rates. Although total mean DER can be considered low, there are differences between dyads, as already implied by higher error rates for same sex dyads than different sex dyads, $t(61) = 4.16$, $p = 1.01e\text{-}04$. While the error produced through silence detection and voice activity detection (MSp + FASp) seems to show high stability throughout dyads (Mean: 3.11, Std: 1.27), SpE is more prone to vary over dyads (Mean: 2.50, Std: 2.12). This is confirmed by the correlation of the total DER and the SpE, $r(105) = 0.83$, $p = 1.24e\text{-}28$. This implies that the variability in total DER is mainly produced by the SpE. Forty of 107 dyads presented a total DER below 5%. Ninety-one of 107 dyads had a total DER below 7.5%. Five dyads showed total DER above 10%. Hence there are dyads for which the method had somewhat increased DER (16 dyads with DER above 7.5%). Mean FASp error rates are below 1%, mean MSp error rates are located just above 2%.

As expected, the dyadic out-of-bag error did correlate positively with the dyadic SpE, providing evidence for the usefulness of the out-of-bag error to estimate the quality of the diarization for specific dyads ($r(105) = 0.85$, $p = 1.65e\text{-}30$, see **Figure 1**).

**TABLE 1 |** Mean *error rates* (Std) in percent over all dyads ($n$ = 107), same-sex dyads ($n$ = 44), and different-sex dyads ($n$ = 63).

| | Total DER | Speaker Error | MSp | FASp |
|---|---|---|---|---|
| All dyads | 5.61 (2.19) | 2.50 (2.12) | 2.60 (1.07) | 0.51 (0.88) |
| Same-sex dyads | 6.48 (2.57) | 3.60 (2.58) | 2.47 (0.98) | 0.41 (0.75) |
| Different-sex dyads | 5.01 (1.64) | 1.74 (1.28) | 2.69 (1.13) | 0.57 (1.04) |

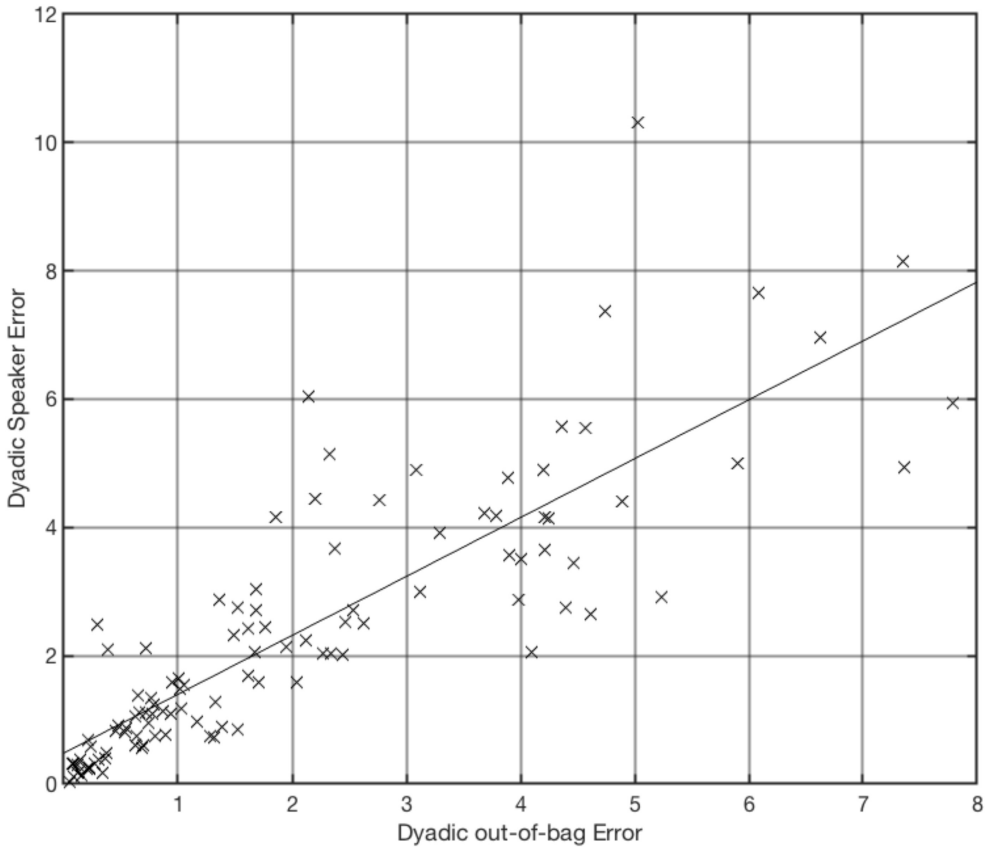


**FIGURE 1 |** Dyadic speaker error (e.g., speaker-1 predicted instead of speaker-2 and vice versa) in percent against dyadic out-of-bag error over all dyads ($n$ = 107), $r(105) = 0.85$, $p$ = 1.65e-30.

## DISCUSSION

Speaker diarization in psychotherapy research has both been performed in a manual, time-consuming (Imel et al., 2014; Reich et al., 2014; Soma et al., 2019), and, alternatively, unsupervised, automated fashion (Xiao et al., 2015; Nasir et al., 2017a,b). For certain scientific contexts, a supervised procedure can be favorable, as it greatly reduces effort (manpower, time, and costs) compared to manual diarization and yields low error rates. In this study, we have described a method for supervised speaker diarization feasible for the dyadic nature of talk therapy. The method requires that the user manually creates a learning set of approximately 5 min cumulative length per speaker. A random forest classifier is trained from the learning set, one for each dyad, using speech features extracted by the MATLAB Audio Analysis Library (Giannakopoulos and Pikrakis, 2014). The classifier is then set out to diarize the whole amount of data (sessions) of this respective dyad. The distinction between voiced and unvoiced windows is made using an already existing procedure for voice activity detection by Segbroeck et al. (2013) and a custom silence detection algorithm. The method is made publicly available (Fürer, 2020). A major advantage of the study is that an open source speech corpus was used to present first results of the proposed method. The availability of the corpus allows other researchers to present results of other methods on the same data set or allows the test of the impact of improvements to the here proposed method.

The method shows satisfying diarization error rates (Mean: 5.61, Std: 2.19), comparable to other fully supervised methods (Zhang et al., 2019). Error rates are higher for same sex dyads (Mean: 6.48, Std: 2.57) than for different sex dyads (Mean: 5.01, Std: 1.64), $t(61) = 4.16$, $p$ = 1.01e-04. This result is expected. The classifier is faced with features of higher degree in similarity when dealing with same-sex dyads, resulting in higher error rates. The difference of diarization error in those groups is mainly due to speaker error (confusion of the classifier toward the distinction of speaker one and speaker two). Speaker error and the total diarization error correlate with $r(105) = 0.83$, $p$ = 1.24e-28, indicating that the total diarization error is mainly produced by speaker error, while missed speech and false alarm speech errors are more stable across dyads. While false alarm speech rates are substantially low (below 1%), missed speech rates are located around 2.5%. Low false alarm speech rates reflect the additional use of silence detection, which has shown to be very robust in differentiating silence windows from non-silence windows. *Post hoc* analyses for silence detection over the whole test set (all dyads together) reveal a miss rate (non-silence predicted instead of

silence) of 0.70% and a false alarm rate (silence predicted instead of non-silence) of 1.51%. Considering the synthetic nature of the corpus used in this study, where the audio files of the corpus contain only silence or speech (no noise), voice activity detection may seem needless besides silence detection. In an environment, where one can be sure that no noise occurs (only speech or silence), the sole use of silence detection can be considered favorable. For later use of the method on naturalistic data, however, where noise may well be part of the equation, voice activity detection is indispensable and is therefore introduced as well. Both silence detection and voice activity detection have been incorporated in the code published (Fürer, 2020).

For 16 (out of 107) dyads, total diarization error exceeds 7.5%. When working with real psychotherapy data, it would be practical to be able to identify these dyads without creating a separate test set. Therefore, we tested whether the out-of-bag error presents a good estimate for the dyadic speaker error. The correlation showed to be high, $r(105) = 0.85$, $p = 1.65e\text{-}30$. We argue that the out-of-bag error can be used to make assumptions toward the quality of diarization, maybe leading to the exclusion of specific dyads, for reasons of error management. It is encouraged for future research to include the out-of-bag error as moderator variable to control for noise.

## LIMITATIONS AND CHALLENGES

In comparison to manual annotations of speaker identity, unsupervised and supervised procedures of speaker diarization will be error prone. It is therefore important for future studies of this realm to report how and to what extent diarization errors influence the research findings at hand. As we reported, the out-of-bag error can be used for this purpose. However, there are no clear guidelines, for example, indicating the need to exclude a dyad for reasons of untolerable diarization error. Consequently, researchers are encouraged to at least publish diarization error rates and to test whether study results correlate with the diarization errors found.

Further, applying machine learning methods to psychotherapeutic data involves experience in programming. Proximity to data scientific or machine learning colleagues is not always guaranteed for workgroups invested in psychotherapy research. It was therefore important to us to publish the code used in this study (Fürer, 2020). The procedure is summarized to one function and an extensive explanation of preprocessing steps is given, in order to make it applicable by users with minimal coding experience.

While using a speech corpus may allow testing future improvements, future studies should invest in testing the proposed procedure on real psychotherapy data in order to clarify concerns toward the validity of results. For an application example of the procedure we refer the reader to the study of Zimmermann et al. (2020), which is using the method presented here to analyze the impact of silence across speaker switching patterns in psychotherapy sessions. Dyadic out-of-bag errors were comparable to the errors found here (Mean: 5.3%, Std = 3.3).

In light of the growing interest in interpersonal processes in psychotherapy, the supervised diarization applied in the study at hand may facilitate the exploration of dyadic vocal and conversational processes that may be linked to change processes, treatment outcome, diagnoses, and patient characteristics. Also, it may facilitate process research to uncover trajectories of variables of interest based on audio recordings. By catalyzing studies concerned with speech and conversational measures, psychotherapy research will gain in rater-independent, objective measures that can widely be used by various research groups and thereby provide results that are comparable and reproducible.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## AUTHOR CONTRIBUTIONS

LF, RZ, and VR contributed to the conception, method, and design of the study. LF performed the analysis and wrote the manuscript with support and revisions of NS, MS, KS, and RZ. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This work was supported by the University of Basel and the Psychiatric Hospitals of the University of Basel.

## ACKNOWLEDGMENTS

This study is performed as part of the EAST psychotherapy study [clinicaltrials.gov: NCT02518906; Zimmermann et al. (2018)].

## REFERENCES

Altmann, U. (2013). *Synchronisation Nonverbalen Verhaltens: Weiterentwicklung und Anwendung Zeitreihenanalytischer Identifikationsverfahren*. Berlin: Springer.

Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G., and Vinyals, O. (2012). Speaker diarization: a review of recent research. *IEEE Trans. Audio Speech Lang. Proc.* 20, 356–370. doi: 10.1109/TASL.2011.2125954

Aveline, M. (1992). The use of audio and videotape recordings of therapy sessions in the supervision and practice of dynamic psychotherapy. *Br. J. Psychother.* 8, 347–358. doi: 10.1111/j.1752-0118.1992.tb01198.x

Barras, C., Zhu, X., Meignier, S., and Gauvain, J.-L. (2006). Multistage speaker diarization of broadcast news. *IEEE Trans. Audio Speech Lang. Proc.* 14, 1505–1512. doi: 10.1109/TASL.2006.878261

Barrios, L., Oldrati, P., Santini, S., and Lutterotti, A. (2019). "Evaluating the accuracy of heart rate sensors based on photoplethysmography for in-the-wild analysis," in *Proceedings of the 13th EAI International Conference*

*on Pervasive Computing Technologies for Healthcare - Pervasive Health*, New York, NY.

Bernieri, F. J., and Rosenthal, R. (1991). "Interpersonal coordination: behavior matching and interactional synchrony," in *Fundamentals of Nonverbal Behavior*, eds R. Feldman and B. Rimé (New York, NY: Cambridge University Press), 401–432.

Bone, D., Lee, C.-C., Potamianos, A., and Narayanan, S. (2014). *An Investigation of Vocal Arousal Dynamics in Child-Psychologist Interactions using Synchrony Measures and a Conversation-based Model*. Shanghai: INTERSPEECH.

Breiman, L. (1996). *Out-Of-Bag Estimation*.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Chen, H.-Y., Liao, Y.-H., Jan, H.-T., Kuo, L.-W., and Lee, C.-C. (2016). "A Gaussian mixture regression approach toward modeling the affective dynamics between acoustically-derived vocal arousal score (VC-AS) and internal brain fMRI bold signal response," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. (2012). Interpersonal synchrony: a survey of evaluation methods across disciplines. *IEEE Trans. Affect. Comput.* 3, 349–365. doi: 10.1109/T-AFFC.2012.12

Edwards, E., Brenndoerfer, M., Robinson, A., Sadoughi, N., Finley, G. P., Korenevsky, M., et al. (2018). "A free synthetic corpus for speaker diarization research," in *Speech and Computer*, eds A. Karpov, O. Jokisch, and R. Potapova (Cham: Springer), 113–122. doi: 10.1007/978-3-319-99579-3_13

Friedland, G., Vinyals, O., Huang, Y., and Muller, C. (2009). Prosodic and other long-term features for speaker diarization. *IEEE Trans. Audio Speech Lang. Proc.* 17, 985–993. doi: 10.1109/TASL.2009.2015089

Fürer, L. (2020). *Supervised Dyadic Speaker Diarization (Version v1) [Computer software]*. San Francisco: GitHub.

Galbusera, L., Finn, M. T. M., Tschacher, W., and Kyselo, M. (2019). Interpersonal synchrony feels good but impedes self-regulation of affect. *Sci. Rep.* 9:14691. doi: 10.1038/s41598-019-50960-0

Giannakopoulos, T., and Pikrakis, A. (2014). *Introduction To Audio Analysis: A MATLAB Approach*, 1st Edn, Cambridge, MA: Academic Press.

Goldstein, E. B. (2010). "Sound, the auditory system, and pitch perception," in *Sensation and Perception*, eds J.-D. Hague, and J. A. Perkins, 8th Edn (Wadsworth CENGAGE Learning), 490.

Husain, W., Xin, L. K., Rashid, N. A., and Jothi, N. (2016). "Predicting Generalized Anxiety Disorder among women using random forest approach," in *Proceedings of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur.

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., et al. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *J. Counsel. Psychol.* 61, 146–153. doi: 10.1037/a0034943

Imel, Z. E., Steyvers, M., and Atkins, D. C. (2015). Computational psychotherapy research: scaling up the evaluation of patient-provider interactions. *Psychotherapy* 52, 19–30. doi: 10.1037/a0036841

Kathania, H. K., Shahnawazuddin, S., Ahmad, W., and Adiga, N. (2019). "On the role of linear, mel and inverse-mel filterbank in the context of automatic speech recognition," in *Proceedings of the 2019 National Conference on Communications (NCC)*, Bangalore.

Kleinbub, J. R. (2017). State of the art of interpersonal physiology in psychotherapy: a systematic review. *Front. Psychol.* 8:2053. doi: 10.3389/fpsyg.2017.02053

Kleinbub, J. R., and Ramseyer, F. (2019). *RMEA: Synchrony in Motion Energy Analysis (MEA) Time-Series (Version 1.1.0) [Computer Software]*. Available online at: https://cran.r-project.org/web/packages/rMEA/index.html (accessed July 17, 2020).

Koole, S. L., and Tschacher, W. (2016). Synchrony in psychotherapy: a review and an integrative framework for the therapeutic alliance. *Front. Psychol.* 7:862. doi: 10.3389/fpsyg.2016.00862

Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artif. Intellig. Rev.* 39, 261–283. doi: 10.1007/s10462-011-9272-4

Lord, S. P., Sheng, E., Imel, Z. E., Baer, J., and Atkins, D. C. (2015). More than reflections: empathy in motivational interviewing includes language style synchrony between therapist and client. *Behav. Ther.* 46, 296–303. doi: 10.1016/j.beth.2014.11.002

Marci, C. D., Ham, J., Moran, E., and Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *J. Nerv. Ment. Dis.* 195, 103–111. doi: 10.1097/01.nmd.0000253731.71025.fc

Masías, V. H., Krause, M., Valdés, N., Pérez, J. C., and Laengle, S. (2015). Using decision trees to characterize verbal communication during change and stuck episodes in the therapeutic process. *Front. Psychol.* 6:379. doi: 10.3389/fpsyg.2015.00379

Nasir, M. D., Baucom, B. R., Bryan, C. J., Narayanan, S. S., and Georgiou, P. (2017a). Complexity in speech and its relation to emotional bond in therapist-patient interactions during suicide risk assessment interviews. *Interspeech* 2017, 3296–3300. doi: 10.21437/Interspeech.2017-1641

Nasir, M. D., Baucom, B. R., Georgiou, P., and Narayanan, S. (2017b). Predicting couple therapy outcomes based on speech acoustic features. *PLoS One* 12:e0185123. doi: 10.1371/journal.pone.0185123

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., and Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Front. Psychol.* 10:2970. doi: 10.3389/fpsyg.2019.02970

Panayotov, V., Chen, G., Povey, D., and Khudanpur, S. (2015). "Librispeech: An ASR corpus based on public domain audio books," in *Proceedings of the2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., and Kirschner, P. A. (2016). "Investigating collaborative learning success with physiological coupling indices based on electrodermal activity," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK*, Amsterdam.

Ramseyer, F. (2013). "Synchronized movement in social interaction," in *Proceedings of the 2013 Inputs-Outputs Conference on An Interdisciplinary Conference on Engagement in HCI and Performance - Inputs*, New York, NY.

Ramseyer, F. (2019). Motion Energy Analysis (MEA). A primer on the assessment of motion from video. *J. Counsel. Psychol.* 67:536.

Ramseyer, F., and Tschacher, W. (2006). Synchrony: a core concept for a constructivist approach to psychotherapy. *Construct. Hum. Sci.* 11, 150–171.

Ramseyer, F., and Tschacher, W. (2008). "Synchrony in dyadic psychotherapy sessions," in *Simultaneity*, eds S. Vrobel, O. E. Rössler, and T. Marks-Tarlow (Singapore: World Scientific), 329–347. doi: 10.1142/9789812792426_0020

Ramseyer, F., and Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *J. Consult. Clin. Psychol.* 79, 284–295. doi: 10.1037/a0023419

Ramseyer, F., and Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* 5:979. doi: 10.3389/fpsyg.2014.00979

Reich, C. M., Berman, J. S., Dale, R., and Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *J. Soc. Clin. Psychol.* 33, 481–494. doi: 10.1521/jscp.2014.33.5.481

Reynolds, D. A., and Torres-Carrasquillo, P. (2005). "Approaches and applications of audio diarization," in *Proceedings of the (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA.

Rubel, J. A., Zilcha-Mano, S., Giesemann, J., Prinz, J., and Lutz, W. (2020). Predicting personalized process-outcome associations in psychotherapy using machine learning approaches—A demonstration. *Psychother. Res.* 30, 300–309. doi: 10.1080/10503307.2019.1597994

Segbroeck, M. V., Tsiartas, A., and Narayanan, S. (2013). *A Robust Front end for VAD: Exploiting Contextual, Discriminative and Spectral Cues of Human Voice. INTERSPEECH*. Shanghai: Shanghai International Convention Center.

Sinclair, M., and King, S. (2013). "Where are the challenges in speaker diarization?," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC.

Slaney, M. (1998). *Auditory Toolbox, Version 2. Technical Report*. Palo Alto, CA: Interval Research Corproation.

Soma, C. S., Baucom, B. R. W., Xiao, B., Butner, J. E., Hilpert, P., Narayanan, S., et al. (2019). Coregulation of therapist and client emotion during psychotherapy. *Psychother. Res.* 30, 591–603. doi: 10.1080/10503307.2019.1661541

Sun, B., Zhang, Y., He, J., Yu, L., Xu, Q., Li, D., et al. (2017). "A random forest regression method with selected-text feature for depression assessment," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge - AVEC*, Cham.

Theodoros, G., and Aggelos, P. (2014). *Introduction to Audio Analysis—1st Edn*. Available online at: https://www.elsevier.com/books/introduction-to-audio-analysis/giannakopoulos/978-0-08-099388-1 (accessed July 17, 2020).

Tomicic, A., Pérez, J. C., Martínez, C., and Rodríguez, E. (2017). Vocalization-silence dynamic patterns: a system for measuring coordination in psychotherapeutic dyadic conversations. *Rev. Latinoam. Psicol.* 49, 48–60. doi: 10.1016/j.rlp.2016.09.004

Tranter, S. E., and Reynolds, D. A. (2006). An overview of automatic speaker diarization systems. *IEEE Trans. Audio Speech Lang. Proc.* 14, 1557–1565. doi: 10.1109/TASL.2006.878256

Tschacher, W., Ramseyer, F., and Koole, S. L. (2018). Sharing the now in the social present: duration of nonverbal synchrony is linked with personality. *J. Pers.* 86, 129–138. doi: 10.1111/jopy.12298

Umesh, S., Cohen, L., and Nelson, D. (1999). "Fitting the mel scale," in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, AZ.

Wallert, J., Gustafson, E., Held, C., Madison, G., Norlund, F., von Essen, L., et al. (2018). Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the U-CARE heart randomized controlled trial. *J. Med. Internet Res.* 20:e10754. doi: 10.2196/10754

Weippert, M., Kumar, M., Kreuzfeld, S., Arndt, D., Rieger, A., and Stoll, R. (2010). Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *Eur. J. Appl. Physiol.* 109, 779–786. doi: 10.1007/s00421-010-1415-9

Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C., and Narayanan, S. S. (2015). "Rate my therapist": automated detection of empathy in drug and alcohol counseling via speech and language processing. *PLoS One* 10:e0143055. doi: 10.1371/journal.pone.0143055

Zhang, A., Wang, Q., Zhu, Z., Paisley, J., and Wang, C. (2019). "Fully supervised speaker diarization," in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton.

Zhou, X., Garcia-Romero, D., Duraiswami, R., Espy-Wilson, C., and Shamma, S. (2011). "Linear versus mel frequency cepstral coefficients for speaker recognition," in *Proceedings of the 2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, Waikoloa, HI.

Zilcha-Mano, S. (2019). Major developments in methods addressing for whom psychotherapy may work and why. *Psychother. Res.* 29, 693–708. doi: 10.1080/10503307.2018.1429691

Zimmermann, R., Fürer, L., Schenk, N., Koenig, J., Roth, V., Schlüter-Müller, S., et al. (2020). Silence in the psychotherapy of adolescents with borderline personality pathology. *Pers. Disord. Theor. Res. Treat.* doi: 10.1037/per0000402

Zimmermann, R., Krause, M., Weise, S., Schenk, N., Fürer, L., Schrobildgen, C., et al. (2018). A design for process-outcome psychotherapy research in adolescents with borderline personality pathology. *Contemp. Clin. Trials Commun.* 12, 182–191. doi: 10.1016/j.conctc.2018.10.007

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**S2: Silence in the psychotherapy of adolescents with borderline personality pathology**

The second paper can be regarded as a natural extension of the devised method, providing first results on the simplest variable of interpersonal coordination devisable from the conversational process: silence. Silence is a co-production of patient and the therapist and happens either between same speaker speaking turns or between speaker switches (Cuttler et al., 2019). AIT advises therapists to hold an active stance and discourages the use of silence as a therapeutic technique (Foelsch et al., 2014). We assumed that most silences are unintentional. In manner of before mentioned relational assumptions, we hypothesized that silence co-occurs in sessions where the ongoing mutual communication and regulation of self-states is resinous. We proposed that silence embodies a decoupling of the patient and therapist's cognitive dyadic self-organization and indicates a shift to more private thought processes of the interlocutors (Laroche et al., 2014). While rhythmic patterns facilitate information processing and interpersonal predictions (Jaffe et al., 2001), silence episodes accompany a halt in the information flow between patient and therapist. Phenomenologically, they are also part of the withdrawal rupture code "minimal response", which was rated with the highest frequency in this study sample (Schenk et al., 2019, 2020).

Further we hypothesized that the smoothness of the dyadic exchange is hindered by higher deficits in personality functioning, and we validated the session-wise amount of silence using patient ratings of smoothness (Foubert, 2017, 2018; Foubert et al., 2020). In a more clinical hypothesis, we tested the clinical stance that silences as a therapeutic technique should be depreciated in adolescent patients with BPD, by predicting patient ratings of the goodness of the session with the amount of silence.

# Silence in the psychotherapy of adolescents with borderline personality pathology.

**Author:**

Zimmermann, Ronan; Fürer, Lukas; Schenk, Nathalie; Koenig, Julian; Roth, Volker; Schlüter-Müller, Susanne; Kaess, Michael; Schmeck, Klaus

**Publication:**

Personality Disorders: Theory, Research, and Treatment

**Publisher:**

American Psychological Association

**Date:**

Apr 1, 2020



- PMID: **32324008**
- DOI: 10.1037/per0000402

# Silence in the Psychotherapy of Adolescents with Borderline Personality Pathology

Ronan Zimmermann[1,2], Lukas Fürer[1], Nathalie Schenk[1], Julian Koenig[3,4], Volker Roth[6], Susanne Schlüter-Müller[1], Michael Kaess[3,5], Klaus Schmeck[1]

[1]*Child and Adolescent Psychiatric Research Department, Psychiatric University Hospital, University of Basel, Switzerland*

[2]*Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel, Switzerland*

[3]*University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Switzerland*

[4]*Section for Experimental Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre of Psychosocial Medicine University of Heidelberg, Germany*

[5]*Section for Translational Psychobiology in Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre of Psychosocial Medicine University of Heidelberg, Germany*

[6]*Department of Mathematics and Computer Science, University of Basel, Switzerland*

**Corresponding author:**

Ronan Zimmermann
Child and Adolescent Psychiatric Hospital (UPK)
Wilhelm Klein-Strasse 27
4002 Basel, Switzerland
ronan.zimmermann@unibas.ch

## Abstract

Silence in psychotherapy has been associated with different, sometimes opposing meanings. This study investigated silence during adolescent identity treatment (AIT) in adolescent patients with borderline personality pathology. A more active therapeutic approach with less silence is advised in AIT. It was hypothesized that a session with more silence might be negatively perceived by adolescent patients. A total of 382 sessions that involved 21 female patients were analyzed. Silence was automatically detected from audio recordings. Diarization (segmenting an audio according to speaker identity) was performed. The patient's perception of the sessions was measured with the Session Evaluation Questionnaire. The amount of silence in the different speaker-switching patterns was not independent of one other. This finding supports the hypothesis of mutual attunement of patient and therapist concerning the amount of silence in a given session. Sessions with less silence were rated as being both smoother and better. The potential implications for clinical practice are discussed. The investigation of turn-taking and interpersonal temporal dynamics is relevant for psychotherapy research. The topic can be addressed efficiently using automated procedures.


*Keywords:* silence, session impact, personality pathology, adolescence, psychotherapy

# Introduction

Silence during psychotherapy and, generally, during conversation can be defined as the meaningful absence of speech. It appears as transitional gaps between speakers or as pauses between units of the same speaker (Heldner & Edlund, 2010). Psychotherapy can be described as a *specialized speech exchange system* in which the conversation takes characteristic forms in the pursuit of a specific effect (Levinson, 2016).

In psychotherapy, silence is sometimes associated with client resistance or withdrawal (Daniel, Folke, Lunn, Gondan, & Poulsen, 2018). In contrast, silence is also considered to be one of the psychotherapist's tools (Buchholz, 2018; Ladany, Hill, Thompson, & O'Brien, 2006). In line with previous impressions, Levitt (2001) showed that the scientific literature associates silence in psychotherapy with various meanings and that there is no homogeneous understanding of silence: Silence in the psychotherapeutic process can be obstructive, productive, or neutral.

It is important to diagnose borderline personality disorder (BPD) already in adolescence (Chanen, Sharp, Hoffman, & Global Alliance for Prevention and Early Intervention for Borderline Personality Disorder, 2017). Subthreshold BPD in adolescents is associated with comparable deficits as full-blown BPD in adolescents in terms of psychopathological distress and the impact on health-related quality of life (Kaess, Fischer-Waldschmidt, Resch, & Koenig, 2017). The term borderline personality pathology (BPP) will subsequently be used to designate both these populations (subthreshold and full-blown BPD).

Early BPP treatment is important to limit deficits in psychosocial functioning and its consequences for the patients' personal development (Chanen, Jovev, & Jackson, 2007; Chanen, Jovev, McCutcheon, Jackson, & McGorry, 2008; Chanen et al., 2009; Miller,

Muehlenkamp, & Jacobson, 2008). Furthermore, experts broadly agree on the necessity of specialized interventions.

At least four manualized therapeutic approaches for adolescents with BPP are currently available: dialectical behavior therapy for adolescents (DBT-A), mentalization-based treatment for adolescents (MBT-A), adolescent identity treatment (AIT), and schema-focused psychotherapy for adolescents (SFT-A; Fleischhaker et al., 2011; Foelsch et al., 2014; Loose, 2015; Rossouw & Fonagy, 2012). AIT is a psychodynamic approach for the treatment of personality disorders in adolescents. It integrates modified elements of transference-focused psychotherapy (TFP: Kernberg, Yeomans, Clarkin, & Kenneth, 2008) with psychoeducation, behavior-oriented home plans, and a stronger focus on working with parents and family. As in TFP, the main techniques of AIT are clarification, confrontation, and interpretation, with emphasis on affects in the here and now, as well as the dominant object-relationship dyads. During the psychoeducation portion of the treatment, both patients and parents are informed about the etiology and course of symptoms, as well as the specificities of relationship building and maintenance, autonomy, limit setting, and affect regulation in the adolescent BPD patients. A written home-plan organizes the overt behavioral interactions between the adolescent and his/her family, provides rewards and consequences for behavior, and clarifies discrepancies in perception between the adolescents and their family.

One of the adaptations of AIT to its targeted adolescent population is the emphasis on an active therapeutic style, with a more careful and reduced use of silence by the psychotherapist compared to the treatment of adults. This adaptation is due to the idea that longer silences are counterproductive in the treatment of adolescents. This view can be understood considering that the conversational styles of adults and adolescents differ: adolescents tend to use faster speech rates and less pausing (Beaumont, 1995; Beaumont & Wagner, 2004).

Gensler (2015) describes that it might not be natural for adolescents to talk and open up to their parents or their therapists, their answers are usually brief, and it is difficult to get into a conversation. The therapist is often required to work hard to get the conversation running. In this situation, an attitude of genuine interest and curiosity toward the adolescent (which is emphasized by AIT and other manualized therapeutic approaches for adolescents with BPP, for example DBT-A or MBT-A) is described as helpful. This problem might be accentuated in adolescents with BPP because BPD affects interpersonal relationships as well as the therapeutic work, in the sense that building trust and stable relationships is problematic (McCutcheon, Chanen, Fraser, Drew, & Brewer, 2007; McMain, Boritz, & Leybman, 2015; Sansone & Sansone, 2013). The current study investigated the occurrence of silence during psychotherapy sessions and its association with the rating of the respective session by the patients. We hypothesized that more silence is associated with a more negative perception of the session. The perception of the session was measured with the dimensions “smoothness” and “goodness” of the Session Evaluation Questionnaire (SEQ: Stiles et al., 1994).

The scientific literature provides vast evidence that depressive symptoms are correlated to a lower speech pace and more silences (Cummins et al., 2015; Ellgring & Scherer, 1996;). Therefore, we considered the level of depression as a potentially confounding factor. Furthermore, we considered the patients’ level of personality functioning as a potentially confounding factor.

Based on previous research in the context of psychotherapy, silence was defined as the absence of speech for three or more seconds (Daniel et al., 2018; Frankel et al., 2006; Stringer et al., 2010). One must consider that silences in a dyadic setting are bound to occur in one of four possible speaker-switching patterns:

Therapist speaking turn – silence – patient speaking turn (T_P);

Patient speaking turn – silence – therapist speaking turn (P_T);

Therapist pauses and continues speaking after the silence (T_T);

Patient pause (P_P).

Silence is a co-production between the patient and the therapist (i.e. any speaker could in theory speak at any time). It has been shown that interlocutors tend to match the duration of pauses (Jaffe et al., 2001). Therefore, it is expected that silence can be evaluated as an overall phenomenon for the purpose of this study. Given that the manual detection of silence is very work intense (Daniel et al., 2018; Frankel et al., 2006), the current study employed automatic silence detection.

## Method

This study is part of the multi-center study 'Evaluation of Adolescent Identity Treatment' (Schmeck et al., 2018; Zimmermann et al., 2018) that has been registered at clinicaltrials.gov. (NCT02518906). The current analyses are based on the entire available data collected at one participating center (Psychiatric Hospitals of the University of Basel). Ethical approval was obtained from the local ethics committee. All adolescents, their parents, and the therapists provided written informed consent.

### Sample

The sample comprised 382 psychotherapeutic sessions that were planned to last for approximately 50 minutes each. They stemmed from 21 AIT psychotherapies of female adolescent patients with BPP. The following inclusion criteria were applied for the patients: age 13–19 years; three or more BPD criteria (Structured Clinical Interview for DSM-IV Axis II Personality Disorders [SCID-II]; First, Gibbon, Spitzer, Williams, & Benjamin, 1997); and identity diffusion according to the Assessment of Identity Development in Adolescence (AIDA;

total *t* score > 60; Goth et al., 2012; Lind, Vanwoerden, Penner, & Sharp, 2019). Although the recruitment was open to male and female patients, only two male patients were enrolled. Consequently, the male patients were excluded from the current analyses to increase the homogeneity of the sample. The mean age of the remaining 21 female patients was 16.3 (*SD* = 1.6) years. Fifteen patients presented with BPD and six with sub-threshold-BPD (three or four fulfilled BPD criteria in SKID-II). Six patients dropped out of treatment but were included in this study (one of them continued with a new therapist, but these sessions were not included in the current study). In two patients, there was a therapist change because the therapist switched occupations. For these cases, only the sessions with the first therapists were included (for statistical reasons). Eleven recordings of therapeutic sessions were missing due to technical difficulties or human failure. Figure A in the Appendix shows the available sessions for each patient. Eight psychotherapists were involved in the current study, two of whom were male.

**Assessments and Data Acquisition**

All instruments and their time of administration used in this study are described in detail elsewhere (Zimmermann et al., 2018). The following section will focus on the instruments relevant for the current analyses. At baseline (before the start of psychotherapy), the level of personality functioning was measured with the LoPF-Q 12–18 total score (Goth, Birkhölzer, & Schmeck, 2018). This measure consists of 97 5-point Likert-type items and was constructed specifically for adolescent populations. The LoPF model encompasses four dimensions, two of which are self-related (identity and self-direction), and two are of an interpersonal nature (empathy and intimacy). The dimensions have a pathological and healthy pole and can be summed to a total score that represents the severity of personality pathology in terms of personality functioning with high values indicating severe personality pathology. The total score differentiated those adolescents with personality disorders from a school sample with a large effect size (*d* = 2.1 *SD*).

Depression was evaluated with the Beck Depression Inventory II (BDI-II: Beck, Ward, Mendelson, Mock, & Erbaugh, 1961); the total score was used. Validity, internal, consistency and reliability of the BDI-II have been shown to be good in German-speaking samples (Kühner, Bürger, Keller, & Hautzinger, 2007).

The psychotherapeutic sessions were video and audio recorded using a boundary microphone mounted to the adjacent wall between the patient and therapist—no more than 1 m from both subjects. The recordings were cut to include only the actual psychotherapeutic process from the moment the therapist invited the patient to start the session in a sitting position to the moment the therapist terminated the session. Audio files were extracted from the video files using MATLAB software (MATLAB Release 2017b, 2017).

After each session, the session smoothness and goodness were rated with the SEQ (Stiles et al., 1994). The instructions for this read: "Please circle the appropriate number to show how you feel about this session"; the items are answered on a 7-point Likert type scale. Goodness was evaluated by means of one item that asked the patient to rate the session from good to bad.

**Extraction of Absence-of-Speech Segments**

To extract absence-of-speech segments, a MATLAB-based vocal activity detector frontend was used. The authors of the detector reported equal error rates below 3% (Van Segbroeck, Tsiartas, & Narayanan, 2013). Single and unconnected windows identified as speech were reclassified as silence because these windows likely resulted from noise (Heldner & Edlund, 2010). The output of the voice activity detector system was adapted to 0.2-second windows, a design that preserves speech in case of contradictory predictions in a window.

**Speaker Diarization**

Speaker diarization (determining who is speaking and when; Anguera et al., 2012) was performed based on a semi-automatic procedure. A human scientific assistant extracted

learning material, which was then used to train a machine learning algorithm to perform diarization of the complete material. The learning set was extracted for each dyad from two initial, two middle, and two final sessions. The set comprised a list of start and stop time stamps of samples of 'patient speech' or 'therapist speech' set with Audacity software (Audacity Team, 2018). We did not use transcripts. The learning set amounted to approximately five minutes of voice recordings per speaker. The whole manual procedure took approximately 30 minutes per dyad. The features for machine learning were calculated in non-overlapping 0.2-second windows using the command 'stFeatureExtraction' from the MATLAB Audio Analysis Library (n.d.). This function computes 35 basic audio features for each window (e.g. mean fundamental frequency or mel frequency cepstral coefficients). The features and learning set were then used in a random forest classifier (Breiman, 2001; Liaw & Wiener, 2002). This decision-tree-based method learned to classify 0.2-second windows of patient or therapist speech based on the extracted features. The average out-of-bag error rate for this classification was 5.3% (*SD* = 3.3). Samples of the classification were verified by listening to the recording with the patient's voice panned to the left and the therapist's voice panned to the right headphone loudspeaker. To assess the quality of the performed silence detection and diarization, we performed a test on a free synthetic speech corpus (Edwards et al., 2018), which showed error rates for our method around 5% on average. Comparison with a manually coded 30-minute extract from one of our therapies revealed an inter-rater reliability (Cohen's kappa) of .80 for identification of silence and correct diarization. The silence detection and diarization method is available open source at the software repository github (Fuerer, 2020).

**Data Aggregation and Definition of Silence**

For each session, data from all 0.2-second windows were obtained, indicating whether it contained therapists' speech, patients' speech or absence of speech. Successive windows containing the same type of segment were then aggregated; this procedure retained information about the starting time of the first and the ending time of the last aggregated window. Silence episodes were defined as segments with 3 or more seconds without speech. The percentage of cumulative silence duration relative to total session time was used as the main parameter for the analysis because it combines incidence and duration and is normalized for session length. The resulting data set comprised four values per session which represent the percentage of silence in the four speaker switching patterns (P_T, T_P, P_P, and T_T; see the end of the 'Introduction').

**Data Analyses**

Analyses were performed with R (R Core Team, 2018). Given that we suspected non-independence of silence in the different speaker switching patterns (Jaffe et al., 2001), we used principal component analysis (PCA) to reduce the dimensionality of the four silence variables (P_P, T_T, T_P, and P_T). PCA performs this task by creating new uncorrelated variables that successively maximize variance. The result is defined by the data at hand and not *a priori* (Jolliffe & Cadima, 2016). This analysis helps to increase interpretability with minimal information loss. We used the function 'prcomp' from the R base package, which uses singular value decomposition on a scaled and centred data matrix.

A linear mixed-effects model was used for hypothesis testing (R package 'nlme': Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018). For the fixed effects, the silence data was used as the dependent variable. The session ratings of the patient (smoothness and goodness of the session) were used as predictors. For the random effects, we initially used a three-level

model, with the intercept by the therapist as the main effect and the intercept by patient nested within the therapist (which is the patient-level effect). Thus, level 1 constituted the sessions (repeated measures), level 2 the 'patients within therapists', and level 3 the therapist main effect. We tested the relevance of the random effects using likelihood ratio tests. Non-relevant random effects were dropped from the model. This modification was used for the therapist main effects in all models. Conclusively, only patients (equivalent to a dyad) were used as the grouping factor in the random effects. The parameter alpha was set to .05 for statistical significance. The variables were log transformed, where necessary, to fit the model assumptions. The model assumptions were verified using residual plots and Q–Q plots. The predictors were centered. The SEQ variables were standardized (*z*-transformed) per dyad.

### *Potentially Confounding Factors*

Since we suspected an overlap of depression and personality functioning, we performed PCA on both variables and added the relevant extracted principal components as additional predictors to the mixed-effects model. Dropout status was controlled for by sub-group analyses. The course of silence over the sessions was considered using a first-order autocorrelative structure of the residuals on the patient level.

## Results

### Description of Silence According to Different Speaker-Switching Patterns

In total, 23,314 silence events were detected in the 382 sessions. The descriptive data are shown in Table 1. All sessions contained silence events, but not every session contained silence events of all speaker-switching types. The median silence length was 4 seconds. On average, 70 silence events occurred per session hour, and the mean total duration of silence per session hour was 6 minutes 9 seconds, which corresponds to 10% of the total time. Therapist pauses

(T_T) were over-represented in terms of incidence and duration; T_P, P_T and P_P pauses were more similar in terms of incidence and duration.

Table 1. A Description of the Silence Events in the Adolescent Identity Treatment

| | Description of silence events (s)* | | | | | | Silence aggregated at the session level | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Pattern** | Median | Q1 | Q3 | Max | Mean | *SD* | Silence count per session hour | Total duration per hour (s) | **Duration in % of (session) time** | % sessions without silences |
| T_P | 4.1 | 3.4 | 5.5 | 9.5 | 5 | 2.4 | 12.1 | 63.3 | **1.8** | 2.7 |
| P_T | 4.2 | 3.6 | 5.4 | 8.8 | 5.1 | 2.4 | 16.1 | 87.3 | **2.4** | 1.5 |
| P_P | 4 | 3.4 | 4.7 | 7.7 | 4.7 | 2.1 | 13.8 | 65.3 | **1.8** | 1.6 |
| T_T | 4 | 3.4 | 5 | 10.7 | 5.1 | 2.5 | 28 | 153.3 | **4.3** | 0.3 |
| Total | 4 | 3.4 | 5.3 | 13.7 | 5 | 2.7 | 70 | 369.2 | **10.3** | 0 |

*Note.* * Calculated using a within-session level and then summarized per subject and between subjects to give each subject equal weight; % session duration (bold) is used as the main parameter for hypothesis testing; T_P = therapist speaking turn – silence – patient speaking turn; P_P = patient pause; P_T = patient speaking turn – silence – therapist speaking turn; P_T = patient speaking turn – silence – therapist speaking turn; T_T = therapist pauses and continues speaking after the silence.

**Correlation of Silence with Session Impact and Patient Pathology**

Four principal components (PC1, PC2, PC3, and PC4) extracted from the silence variables (T_P, P_P, T_T, and P_T) explained 52%, 27%, 14%, and 7% of the variance in the silence data, respectively. Table 2 shows the correlation of the original variables with the principal components.

Table 2. Principal Component (PC) Loadings

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| **T_P Silence** | 0.56 | -0.25 | 0.62 | 0.49 |
| **P_P Silence** | 0.46 | -0.63 | -0.21 | -0.59 |
| **P_T Silence** | 0.54 | 0.27 | -0.69 | 0.39 |
| **T_T Silence** | 0.42 | 0.68 | 0.31 | -0.51 |

*Note.* The table shows the correlation of the extracted principal components (PC1, PC2, PC3, and PC4) with the original silence variables (T_P, P_P, P_T, and T_T); T_P = therapist speaking turn – silence – patient speaking turn; P_P = patient pause; P_T = patient speaking turn – silence – therapist speaking turn; P_T = patient speaking turn – silence – therapist speaking turn; T_T = therapist pauses and continues speaking after the silence.

PC1 correlated positively with all four original silence variables (component loadings from .42 to .56). This finding is thus consistent with the idea that silence can be evaluated as an overall phenomenon that is co-produced by the patient and therapist due to mutual attunement of the speakers. High PC1 values involve more silence in all speaker-switching patterns. PC1 will be used (instead of four separate silence variables T_P, P_P, T_T and P_T) for the following analyses concerning the correlation of silence with the perception of the sessions. Notably, PC2 showed a negative correlation between patient and therapist silences, especially regarding patient and therapist pauses. PC3 and PC4 only explained minor portions of the variance.

Concerning the potential confounding variable depression and personality pathology, the PCA revealed a strong PC1 that explained 94% of the variance in both these variables. This component correlated positively with personality pathology (.99) and positively but weakly with depression (.16). This result clearly demonstrated that there was almost no variance of depression in the current sample after considering shared variance with personality functioning. The resulting PC1 of this PCA was used as confounding variable in the mixed effects-model (see below).

**Hypothesis Testing**

The linear mixed-effects model showed that smoothness of the sessions rated by the patients was significantly negatively correlated with the PC1 silence component (b = -.32, *SE* = .06, $t(359) = -5.7$, $p < .001$). This finding indicates that the smoother a session is perceived by the patients, the less silence emerges in all speaker-switching patterns. Similarly—but independent from smoothness—sessions with less silence were also considered better by the patients (b = -.16, *SE* = .00, $t(359) = -2.9$, $p < .004$). Figure 1 shows these correlations in the single patients.

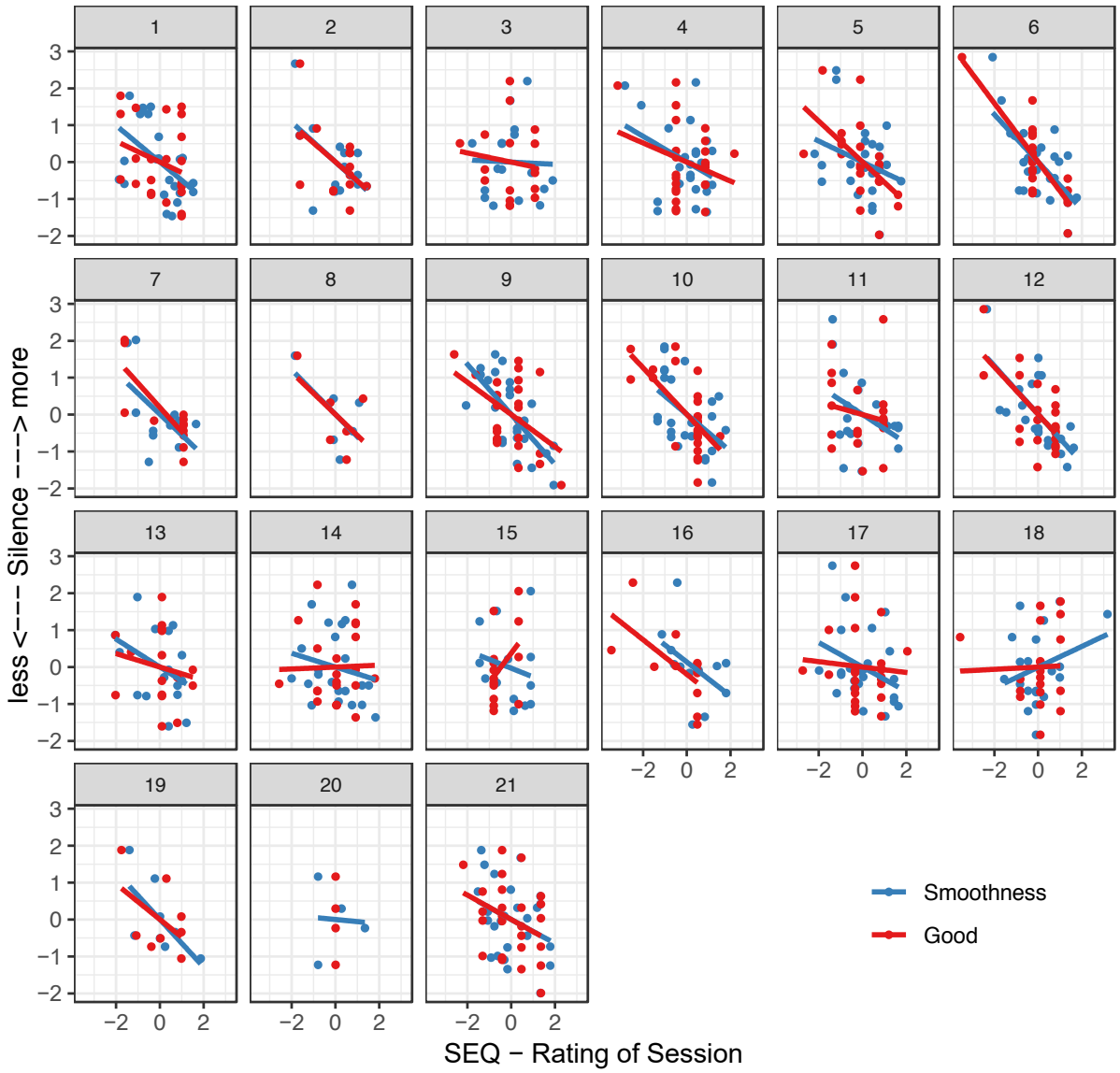

**Figure 1. Correlation of smoothness and goodness with silence**

Each panel (1, 2, 3, …) represents a patient-therapist dyad. Each data point represents a session. In this figure, the session ratings (smoothness and goodness) and the percentage of silence per session are standardised (*z*-scores) at the patient level. SEQ = Session evaluation questionnaire.

The principal component that represented psychopathology was not a significant predictor of silence ($p = .282$). To control for dropout status, the patients who remained in treatments and dropout patients were analyzed separately. For the six dropout cases, the coefficients of smoothness (b = -.35) and goodness (b = -.15) were similar to the overall sample. The same was true for the patients remaining in treatment (b of smoothness = -.31, b of goodness = -.16).

## Discussion

In the present study, silence was measured during AIT of patients with BPP using automatic procedures. For this purpose, silence was defined as the absence of speech for 3 or more seconds. The results showed that silence episodes occurred in all psychotherapy sessions; on average, they comprised 10% of the total session time. The average maximum silence episode in a session was 14 seconds in length, which might seem rather short for a psychotherapeutic setting (comparison with other samples is required) and might be related to the adolescent population's conversational style and the active therapeutic stance recommended in AIT. We hypothesized that sessions with more silence would be perceived as more negative by the patients. This hypothesis was confirmed because sessions with more silence were perceived as less smooth; that is, they were rated as more unpleasant, rough, difficult, and uncomfortable by the patients. At the same time and independent from smoothness, the patients perceived the sessions with less silence as better. The association of more silence with less smoothness seems to be particularly robust because it was observable in almost all patients

(Figure 1). The result concerning smoothness is consistent with the idea that silences in the psychotherapy might not be well tolerated by the investigated female adolescent patients with BPP. The additional result concerning goodness of the sessions was somewhat surprising and underscores that sessions with more silence are not appreciated by this population.

Levitt (2001) showed that the scientific literature associates silence in psychotherapy with various meanings; there is no homogenous understanding of silence. Our approach decontextualizes single silence events and sums them into a total amount per session. In this context, silence can be understood as a general tendency comprises different 'types' of silence. Coordination of timing is not a proper aim of daily interaction (nor of psychotherapy) and likely emerges from self-organization of the between-persons relational dynamics (Laroche et al., 2014). We assume that in some cases silences might be introduced intentionally (e.g. by the therapist). However, considering that we are analyzing a large number of silence events (more than 23,000 in this study), we think that most silences are not intentional. We think that the therapist's best leverage on silence would be through his or her psychotherapeutic stance (e.g. curious vs. abstinent). In our perspective, silence signifies a decoupling of the patient and therapist's cognitive processes and a shift from direct sharing to a mode where the interlocutors' cognitive processes become more private. Researchers have suggested that the rhythmic interpersonal relationship is constantly assessed by the interlocutors, with rhythmic patterns facilitating information processing and interpersonal predictions (Byers, 1976; Jaffe et al., 2001). We think that silence decreases the predictability of the interlocutor's current cognitive processes. Whether silence is valued as problematic or unsmooth depends on the assumptions that the interlocutors make about their counterpart in a situation with reduced transparency of the interlocutor's current cognitive processes. According to this perspective, it is possible that silence in psychotherapy of adolescents with BPP is (at least occasionally) useful despite a general tendency of perceiving sessions with more silence as negative. Thus, the result of the

current study does not support the view that silence should be omitted. However, psychotherapists might want to proceed with extra care when introducing silences or when sessions contain a lot of silence. A potential danger might come from the differences of adult and adolescent conversational styles (including pausing behavior, but also concerning other parameters). These differences potentially result in a 'clash in conversational styles' that can lead to misunderstandings based on different assumptions of the speakers (Beaumont & Wagner, 2004) with regard to the significance of silence. To a certain extent, the more active therapeutic stance recommended by AIT is validated by the current findings. The results are also relevant for other manualized treatment approaches for BPP in adolescents (i.e., DBT-A, MBT-A, and SFT-A), which espouse a curious and active therapeutic stance.

It is clear that silence, and even gaps and pauses, convey information to the interlocutors. Longer gaps and pauses are negatively correlated with the perceived willingness and agreement to requests (Roberts, Margutti and Takano 2011), and disagreement and rejection become more likely in normal conversation after longer gaps. Interlocutors are sensitive to relatively small variations, with a turning point at 700 ms in the investigated non-psychotherapeutic context (Levinson & Torreira, 2015). It is possible that patients with BPD are more sensitive to a negative perception of silence: Patients with BPD have a bias toward negative emotions, a phenomenon that leads to difficulties in interpersonal relations (American Psychiatric Association, 2013; Carpenter & Trull, 2012). The negative perception of silences in conversations might be accentuated by this bias. Additionally, hypermentalizing, "a social-cognitive process that involves making assumptions about other people's mental states that go so far beyond observable data that others may struggle to see how they are justified" (Sharp & Vanwoerden, 2015, p. 38), described in these patients, might play a role (Sharp et al., 2013). In an analogous manner to the discrepancies between adult and adolescent conversational styles, this bias toward a negative interpretation of social cues in patients with BPD might introduce a

further danger of silences being misunderstood between the patient and therapist. In view of the difficulty with alliance building in patients with BPD (McCutcheon et al., 2007; McMain et al., 2015; Sansone & Sansone, 2013) and the considerable dropout rates (in current study almost 30%), we think that the therapist should be aware that patients might have a tendency to perceive silence more negatively. To address this difficulty, manualized therapeutic approaches for adolescents with BPP recommend an active therapeutic stance with genuine curiosity toward the patient which does not leave room for the occurrence of longer silence episodes. An alternative approach might be clarification (a technique used in AIT and MBT-A) of the meaning of silence in the dyad by actively enquiring and explaining. In this context, it is interesting to note that pauses made by the therapist (T_T) were overrepresented (twice as frequent and long) in the current analysis. It might be of interest to explain this conversational pattern (which might be an aspect of the psychotherapeutic context) to the patient.

The amount of silence episodes is only one aspect of the complex intersubjective temporalities at work in dyadic systems. We think that these relational temporal dynamics are mostly self-organizing and recommend the article by Laroche et al. (2014) for a deeper understanding. A complex interpersonal temporal dynamic is essential for human communication according to this view. We are convinced that certain exchanges are not possible without the dyadically self-organized mutual attunement to the right temporal dynamics for the specific exchange at hand. While this aspect might not be something that psychotherapists can straightforwardly shape because of the inherent complexity, this temporal dynamic is nevertheless an important and probably underestimated aspect of psychotherapy.

Concerning the intersubjective dynamic of silence, the PCA (see the 'Correlation of Silence with Session Impact and Patient Pathology' section) of the silence variables representing the different speaker switching patterns identified a first principal component (PC1) that was positively correlated with silence in all speaker-switching patterns: This

component explained the majority of the variance in the silence data (52%) and can be interpreted as representing the result of mutual attunement. Such synchronization concerning silence was expected based on previous research of turn taking, mainly in mothers and infants (Jaffe et al., 2001). This observed synchronization might also occur in psychotherapy, making silence one of the synchronizing modalities in psychotherapy (see Gaume et al., 2019; Imel et al., 2014; Kleinbub, 2017; Paulick et al., 2017; Reich et al., 2014; Schultz et al., 2016 for other modalities). Interestingly, the second principal component (PC2), which explained 27% of variance in the silence data, showed a negative correlation mainly between patient and therapist pauses. This finding suggests that the intra-speaker silences (pauses) of the two interlocutors are, by tendency, not long at the same time. These findings of opposite tendencies in the silence data draw an interesting overall picture of the intercorrelation of silence in the different speaker-switching patterns: While there is mainly synchronization in the amount of silence, there is a concomitant balancing pattern of silence in the dyadic system concerning pauses. This finding is supported by similar matching, as well as 'compensatory' patterns found in vocal rhythm coordination (Jaffe et al., 2001).

**Limitations and Future Directions**

A separate investigation of depression and personality pathology in our sample was not appropriate because both patient characteristics were intercorrelated, as shown by the PCA. The study did not find any correlation of silence with patient pathology. One must consider that the investigation was based on a relatively small and homogeneous sample in which all patients were diagnosed with BPP. Additionally, depression was not the primary diagnosis; it was only a comorbidity in this sample. To generate clearer conclusions regarding silence in adolescent psychotherapy and silence in BPP, group comparisons of psychotherapies should be investigated in larger samples. For example, it would be of interest to contrast adolescents and adults, patients with different pathologies and different types of psychotherapeutic

interventions. Another topic that was not explicitly addressed in this study is the temporal dynamic of silence in the course of the psychotherapy. This issue was not evaluated because the sample size and the included dropout patients—who participated in fewer sessions—make the sample inadequate for conclusively addressing this question. The question is of interest but should be addressed in larger and more suited samples.

Notably, when considering further research, our proposed method for silence detection is highly efficient. In this study, the method almost effortlessly captured more than 20,000 silence episodes (compared to manual coding) in 382 therapeutic sessions. The error rates were good, especially given that a standard microphone setup was used (only one microphone).

A comparison with more traditional silence characterization methods might help to understand implications and limitations of the applied automated method: Levitt (2001) developed the pausing inventory categorization system (PICS), which might be one of the most utilized silence characterization methods in psychotherapy research. PICS was developed based on grounded theory analysis and interpersonal process recall interviews. Using a heuristic approach, three categories were formulated: neutral silence, productive silence, and obstructive silence. These categories are observer-based ratings that require contextual verbal and paraverbal cues to be identified (as the silence itself does not contain any information). As described in Daniel et al. (2018), the application of PICS requires multiple steps: 1) timing of pauses by first listening the entire sessions and marking pauses directly on the waveform; 2) rating of speaker switching pattern of each pause; and 3) rating of the silence category. To apply PICS, the coders require training to achieve a common understanding. The currently applied automatic method replicates steps one and two of this approach: 1) The detection of pauses (absence of speech) is achieved by analyzing the waveform with a vocal activity detector algorithm (see the 'Extraction of Absence-of-Speech Segments' section); and 2) the speaker switching pattern of the silence episode is determined using a diarization algorithm (see the

‘Speaker Diarization’ section), which also uses features of the waveform. Both of these steps can be accomplished by algorithms with very good (but not perfect) accuracy. One major difference between humans and machines is the level of control over the context. Humans will usually understand what is being said and can notice when something is not right. The algorithms will only do what they were designed to do. Therefore, it is important that the researchers know the audio material that they are studying and, additionally, have a basic understanding of what the algorithm does. Quality and plausibility checks are strongly recommended. We think that the performance of human coders might currently be better in steps one and two when working with a limited amount of audio material. However, human coders have limited sustained attention (vigilance; Warm, Parasuraman, & Matthews, 2008) and will fluctuate in how accurately they can perform these tasks. Inter-rater reliability for the identification of speaker switching patterns (step 2) is very good but not perfect when performed by human coders (Daniel et al., 2018). Finally, when rating manually, the associated workload is resource intense, a factor that makes it difficult to analyze complete psychotherapies (Frankel et al., 2006). Step 3 of Levitt’s method, the differentiation of neutral, obstructive, and productive silence, cannot be automated. However, provided that steps one and two were performed, human coders could now review the identified silences and rate them accordingly.

While the workload issue presents a handicap for the research context, it offers an insuperable obstacle when aiming to implement a method in clinical practice. Potentially, automatic detection of speech parameters might inform supervision in clinical practice. Whether algorithms can provide clinically useful information is an open research question that was not addressed by the current study. We think that the presented method could provide the therapist with meta data on his or her sessions, e.g. whether the current session contains more or less silence than his or her average session. It could also help the therapist to identify silences of particular length when reviewing his or her sessions.

Silence can be analyzed on different levels (e.g. episode or utterance). Given that the current study analyzed the data on a session level, it is difficult to state the exact meaning of a given speaker-switching pattern is. For example, it is difficult to consistently characterize the meaning of a gap after a therapist and before a patient-speaking turn. It is likely that the meaning of T_P, for example, is not consistent. The summing of these events by session can only provide an idea of an overall effect (which would not as such be reflected in a single event). The problem could be attenuated by correlating silence with significant therapeutic events such as ruptures and resolutions or change moments (Krause et al., 2007; Schenk et al., 2019; Zimmermann et al., 2018), thus narrowing down the meanings of silence episodes to more specific contexts. Additionally, the definition of silence as absence of speech for at least 3 seconds is arbitrary and was based on existing literature (Daniel et al., 2018). It would be of interest to investigate potential shifts in meanings when altering this parameter or to compare different settings with regard to their distribution of gaps and pauses.

## Acknowledgements

We would like to thank the participating patients and psychotherapists for trusting us with the scientific analysis of their psychotherapeutic work. We thank Mariane Krause, Nelson Valdez, Pamela Foelsch, and Andrés Borzutzky for their support in designing and conducting the ‘Evaluation of Adolescent Identity Treatment’ study. Furthermore, we would like to thank the Clinical Trial Unit in Basel for help with ethics approval and Andrea Wiencierz for statistical guidance.

## Funding

This work was supported by the University of Basel and the Psychiatric University Hospitals of the University of Basel

## Disclosure of Interest

Klaus Schmeck and Susanne Schlüter-Müller are co-authors of the Manual of Adolescent Identity Treatment.

## Data Availability Statement

The datasets used during the current study are available on reasonable request. Please direct your request to the corresponding author.

## References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, D.C.: American Psychiatric Association.

Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G., & Vinyals, O. (2012). Speaker diarization: A review of recent research. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(2), 356–370. https://doi.org/10.1109/TASL.2011.2125954

Audacity Team (2018). Audacity ®. Free, open source, cross-platform audio software for multi-track recording and editing [Computer software]. Retrieved from https://www.audacityteam.org/

Beaumont, S. L. (1995). Adolescent girls' conversations with mothers and friends: A matter of style. *Discourse Processes*, *20*(1), 109–132. https://doi.org/10.1080/01638539509544933

Beaumont, S. L., & Wagner, S. L. (2004). Adolescent-parent verbal conflict: The roles of conversational styles and disgust emotions. *Journal of Language and Social Psychology*, *23*(3), 338–368. https://doi.org/10.1177/0261927X04266813

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buchholz, M. B. (2018). Kleine Theorie der Pause: Was in therapeutischen Gesprächen auch eine Rolle spielt. *Psyche*, *72*(02), 91–121. https://doi.org/10.21706/ps-72-2-91

Byers, P. (1976). Biological rhythms as information channels in interpersonal communication behavior. In P. P. G. Bateson & P. H. Klopfer (Eds.), *Perspectives in Ethology, Vol. 2* (pp. 135–164). https://doi.org/10.1007/978-1-4615-7572-6_3

Carpenter, R. W., & Trull, T. J. (2012). Components of emotion dysregulation in borderline personality disorder: A review. *Current Psychiatry Reports*, *15*(1), 335. https://doi.org/10.1007/s11920-012-0335-2

Chanen, A. M., Jovev, M., & Jackson, H. J. (2007). Adaptive functioning and psychiatric symptoms in adolescents with borderline personality disorder. *The Journal of Clinical Psychiatry*, *68*(2), 297–306.

Chanen, A. M., Jovev, M., McCutcheon, L. K., Jackson, H. J., & McGorry, P. D. (2008). Borderline personality disorder in young people and the prospects for prevention and early intervention. *Current Psychiatry Reviews*, *4*(1), 48–57.

Chanen, A. M., Mccutcheon, L. K., Germano, D., Nistico, H., Jackson, H. J., & Mcgorry, P. D. (2009). The HYPE clinic: An early intervention service for borderline personality disorder. *Journal of Psychiatric Practice*, *15*(3), 163–172. https://doi.org/10.1097/01.pra.0000351876.51098.f0

Chanen, A. M., Sharp, C., Hoffman, P., & Global Alliance for Prevention and Early Intervention for Borderline Personality Disorder (2017). Prevention and early intervention for borderline personality disorder: A novel public health priority. *World Psychiatry*, *16*(2), 215–216. https://doi.org/10.1002/wps.20429

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

Daniel, S. I. F., Folke, S., Lunn, S., Gondan, M., & Poulsen, S. (2018). Mind the gap: In-session silences are associated with client attachment insecurity, therapeutic alliance, and treatment outcome. *Psychotherapy Research*, *28*(2), 203–216. https://doi.org/10.1080/10503307.2016.1177673

Edwards, E., Brenndoerfer, M., Robinson, A., Sadoughi, N., Finley, G. P., Korenevsky, M., … Suendermann-Oeft, D. (2018). A free synthetic corpus for speaker diarization research. In A. Karpov, O. Jokisch, & R. Potapova (Eds.), *Speech and Computer* (pp. 113–122). New York, NY: Springer International Publishing.

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*(2), 83–110. https://doi.org/10.1007/BF02253071

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, DC: American Psychiatric Press.

Fleischhaker, C., Böhme, R., Sixt, B., Brück, C., Schneider, C., & Schulz, E. (2011). Dialectical behavioral therapy for adolescents (DBT-A): A clinical trial for patients with suicidal and self-injurious behavior and borderline symptoms with a one-year follow-up. *Child and Adolescent Psychiatry and Mental Health*, *5*(1), 3. https://doi.org/10.1186/1753-2000-5-3

Foelsch, P. A., Schlüter-Müller, S., Odom, A. E., Arena, H. T., Borzutzky H., A., & Schmeck, K. (2014). *Adolescent identity treatment*. https://doi.org/10.1007/978-3-319-06868-8

Frankel, Z., Levitt, H., M Murray, D., Greenberg, L., & Angus, L. (2006). Assessing silent processes in psychotherapy: An empirically derived categorization system and sampling

strategy. *Psychotherapy Research*, *16*, 627–638. https://doi.org/10.1080/10503300600591635

Fuerer, L. (2020). *Supervised-Dyadic-Speaker-Diarization* [MATLAB software]. Retrieved from https://github.com/lukasfuerer/Supervised-Dyadic-Speaker-Diarization (Original work published 2020)

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

Gensler, D. (2015). Silence in adolescent psychotherapy. *Journal of Infant, Child, and Adolescent Psychotherapy*, *14*(2), 188–195. https://doi.org/10.1080/15289168.2015.1032635

Goth, K., Birkhölzer, M., & Schmeck, K. (2018). *LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire) German Version: A self-report questionnaire for measuring personality functioning in adolescence – short description*. Offenbach, Germany: Academic-Tests.

Goth, K., Foelsch, P., Schlüter-Müller, S., Birkhölzer, M., Jung, E., Pick, O., & Schmeck, K. (2012). Assessment of identity development and identity diffusion in adolescence—Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child and Adolescent Psychiatry and Mental Health*, *6*(1), 27. https://doi.org/10.1186/1753-2000-6-27

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Kircher, J. C., Baer, J. S., & 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

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*(2), i–149. https://www.jstor.org/stable/3181589

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kaess, M., Fischer-Waldschmidt, G., Resch, F., & Koenig, J. (2017). Health related quality of life and psychopathological distress in risk taking and self-harming adolescents with full-syndrome, subthreshold and without borderline personality disorder: Rethinking the clinical cut-off? *Borderline Personality Disorder and Emotion Dysregulation*, *4*(1), 7. https://doi.org/10.1186/s40479-017-0058-4

Kernberg, O. F., Yeomans, F. E., Clarkin, J. F., & Kenneth, L. N. (2008). Transference focused psychotherapy: Overview and update. *The International Journal of Psychoanalysis*, *89*(3), 601–620. https://doi.org/10.1111/j.1745-8315.2008.00046.x

Kleinbub, J. R. (2017). State of the art of interpersonal physiology in psychotherapy: A systematic review. *Frontiers in Psychology*, *8*, 2053. https://doi.org/10.3389/fpsyg.2017.02053

Krause, M., de la Parra, G., Arístegui, R., Dagnino, P., Tomicic, A., Valdés, N., … Ben-Dov, P. (2007). The evolution of therapeutic change studied through generic change indicators. *Psychotherapy Research*, *17*(6), 673–689. https://doi.org/10.1080/10503300601158814

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). *Der Nervenarzt*, *78*(6), 651–656. https://doi.org/10.1007/s00115-006-2098-7

Ladany, N., Hill, C. E., Thompson, B. J., & O'Brien, K. M. (2006). Therapist perspectives on using silence in therapy: A qualitative study. *Counselling and Psychotherapy Research*, *4*(1), 80–89. https://doi.org/10.1080/14733140412331384088

Laroche, J., Berardi, A. M., & Brangier, E. (2014). Embodiment of intersubjective time: Relational dynamics as attractors in the temporal coordination of interpersonal behaviors and experiences. *Frontiers in Psychology*, *5*, 1180. https://doi.org/10.3389/fpsyg.2014.01180

Levinson, S. C. (2016). Turn-taking in human communication: Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731

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

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lind, M., Vanwoerden, S., Penner, F., & Sharp, C. (2019). Inpatient adolescents with borderline personality disorder features: Identity diffusion and narrative incoherence. *Personality Disorders: Theory, Research, and Treatment*, *10*(4), 389–393. https://doi.org/10.1037/per0000338

Loose, C. (2015). Die Schematherapie bei adoleszenten Patienten mit Borderline-Störung. *Persönlichkeitsstörungen: Theorie und Therapie*, *19*(1), 44–54.

MATLAB Audio Analysis Library—File Exchange, MATLAB Central [Computer software]. (n.d.). Retrieved 24 July 2018, from https://www.mathworks.com/matlabcentral/fileexchange/45831

MATLAB Release 2017b [Computer software]. (2017). Natick, MA: The MathWorks, Inc.

McCutcheon, L. K., Chanen, A. M., Fraser, R. J., Drew, L., & Brewer, W. (2007). Tips and techniques for engaging and managing the reluctant, resistant or hostile young person. *Med J Aust*, *187*(7 Suppl), S64-7.

McMain, S. F., Boritz, T. Z., & Leybman, M. J. (2015). Common strategies for cultivating a positive therapy relationship in the treatment of borderline personality disorder. *Journal of Psychotherapy Integration*, *25*(1), 20–29. https://doi.org/10.1037/a0038768

Miller, A. L., Muehlenkamp, J. J., & Jacobson, C. M. (2008). Fact or fiction: Diagnosing borderline personality disorder in adolescents. *Clinical Psychology Review*, *28*(6), 969–981. https://doi.org/10.1016/j.cpr.2008.02.004

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.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and Nonlinear Mixed Effects Models [Computer software]. Retrieved from https://CRAN.R-project.org/package=nlme

R Core Team (2018). R: A Language and Environment for Statistical Computing [Computer software]. Retrieved from https://www.R-project.org/

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. https://doi.org/10.1521/jscp.2014.33.5.481

Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*(5), 331–354. https://doi.org/10.1080/0163853X.2011.558002

Rossouw, T. I., & Fonagy, P. (2012). Mentalization-based treatment for self-harm in adolescents: A randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(12), 1304–1313.e3. https://doi.org/10.1016/j.jaac.2012.09.018

Sansone, R. A., & Sansone, L. A. (2013). Responses of mental health clinicians to patients with borderline personality disorder. *Innovations in Clinical Neuroscience*, *10*(5–6), 39–43.

Schenk, N., Zimmermann, R., Fürer, L., Krause, M., Weise, S., Kaess, M., Schlüter-Müller, S., & Schmeck, K. (2019). Trajectories of alliance ruptures in the psychotherapy of adolescents with borderline personality pathology: timing, typology and significance. Research in Psychotherapy: Psychopathology, Process and Outcome, 22(2). https://doi.org/10.4081/ripppo.2019.348Schmeck, K., Pick, O. G., Milidou, M., Schenk, N., Schlüter-Müller, S., & Zimmermann, R. (2018). Früherkennung von Persönlichkeitsstörungen. *PTT - Persönlichkeitsstörungen*, *22*(3), 179–185.

Schultz, B. G., O'Brien, I., Phillips, N., McFarland, D. H., Titone, D., & Palmer, C. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201–1220. https://doi.org/10.1017/S0142716415000545

Sharp, C., Ha, C., Carbone, C., Kim, S., Perry, K., Williams, L., & Fonagy, P. (2013). Hypermentalizing in adolescent inpatients: Treatment effects and association with

borderline traits. *Journal of Personality Disorders*, *27*(1), 3–18. https://doi.org/10.1521/pedi.2013.27.1.3

Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in borderline personality disorder: A model and data. *Journal of Infant, Child, and Adolescent Psychotherapy*, *14*(1), 33–45. https://doi.org/10.1080/15289168.2015.1004890

Stiles, W. B., Reynolds, S., Hardy, G. E., Rees, A., Barkham, M., & Shapiro, D. A. (1994). Evaluation and description of psychotherapy sessions by clients using the Session Evaluation Questionnaire and the Session Impacts Scale. *Journal of Counseling Psychology*, *41*(2), 175–185. https://doi.org/10.1037/0022-0167.41.2.175

Stringer, J. v, Levitt, H. M., Berman, J. S., & Mathews, S. S. (2010). A study of silent disengagement and distressing emotion in psychotherapy. *Psychotherapy Research*, *20*(5), 495–510. https://doi.org/10.1080/10503301003754515

Van Segbroeck, M., Tsiartas, A., & Narayanan, S. (2013). A robust frontend for VAD: Exploiting contextual, discriminative and spectral cues of human voice. *InterSpeech*, 704–708.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors, 50*(3), 433–441. https://10.1518/001872008X312152

Zimmermann, R., Krause, M., Weise, S., Schenk, N., Fürer, L., Schrobildgen, C., … Schmeck, K. (2018). A design for process-outcome psychotherapy research in adolescents with borderline personality pathology. *Contemporary Clinical Trials Communications*, *12*, 182–191. https://doi.org/10.1016/j.conctc.2018.10.007

# Appendix

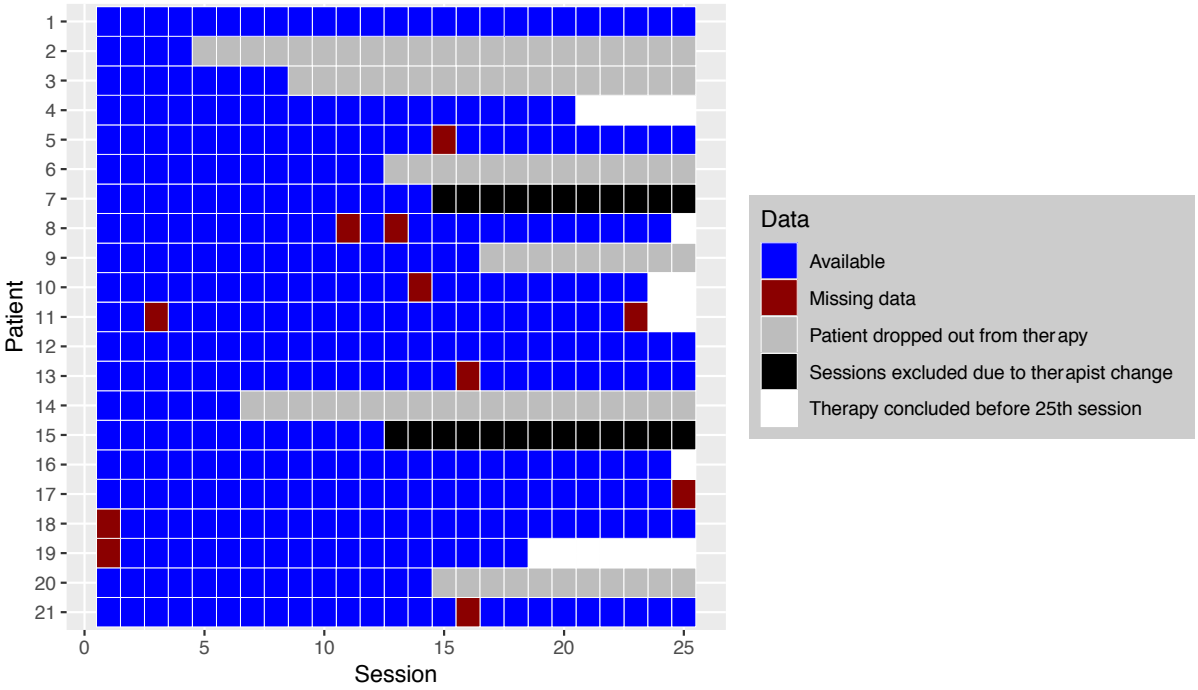


**Figure A. Available data evaluated in this study**

**S3: Machine Learning Facial Emotion Recognition in Psychotherapy Research: Evidence supporting a new technology**

As discussed, momentary activation of affect can lead the way to activation of object relations dyads. In order to identify possible candidate episodes from the flow of therapy, emotion-trajectories of high sampling frequency are needed to describe both patient and therapist in-session affect. The assessment of in-session emotion trajectories for both patient and therapist proves to be a persistent interest when studying the administration and outcome of psychotherapeutic approaches (Greenberg & Safran, 1989). Besides questionnaires and observational approaches (Carryer & Greenberg, 2010), we aim at the assessment of affect during psychotherapy in automated fashion with high resolution. This has been done using video (Arango et al., 2019), audio recordings (Crangle et al., 2019), psychophysiological measurements (Kleinbub, 2017) and text analysis (natural language processing; Halfon et al., 2020). In its crudest form, emotions are described in a two-dimensional space of arousal and valence (circumplex model), arousal indicating the emotional load or level of activation and valence indicating a pleasure dimension (to what extent the emotion are positive or negative; (Posner et al., 2005; Whissell, 2009). Different information streams perform differently in the recognition of these dimensions. For example, the voice offers features for a robust detection of arousal (Bone et al., 2014), however the recognition of specific basic emotions (valence dimension) from vocal features (sentiment analysis: Yadav et al., 2015); emotion recognition: Crangle et al., 2019) is hindered by the fact that the vocal expression of specific emotions differs across speakers (speaker-dependency: Vogt et al., 2008). Given the current state of research, it must be concluded that there is no definite discriminative pattern of speech features indicative of specific basic emotions across speakers (Russell et al., 2003). The human face offers an information stream that is relatively free of this constraint, therefore being a promising new avenue in the measurement of in-session patient and therapist affect. Recent advances in the field of affective computing brought forth open-source libraries for face detection and facial emotion detection using convolutional neural networks (Bradski & Kaehler, 2008; Chollet, 2017). In an extensive programme, we validated such an approach on a large amount of human emotion ratings using the client expressed emotional arousal scale (Carryer & Greenberg, 2010) including both arousal and specific

basic emotion rating. So far, only a handful of studies applied facial machine learning emotion recognition in the context of psychotherapy (Arango et al., 2019; Halfon et al., 2020) and to the best of our knowledge, this is the first study to validate a method of facial emotion recognition in natural psychotherapy data.

**TITLE PAGE**

**Title:** Machine Learning Facial Emotion Recognition in Psychotherapy Research: Evidence supporting a new technology.

**Authors:** Martin Steppan Dr[1], Ronan Zimmermann Dr. [1,2], Lukas Fürer MS[1], Matthew Southward [3] , Julian Koenig[4,5], Michael Kaess Prof. MD [4,6], Johann Roland Kleinbub Dr.[7], Volker Roth Prof. Dr. [8], Klaus Schmeck Prof. MD Dr. [1]

**Author E-mail List:** (separated by commas or returns)
mhsteppan@gmail.com
ronan.zimmermann@unibas.ch
lukas.fuerer@unibas.ch
southward@uky.edu
Julian.Koenig@med.uni-heidelberg.de
michael.kaess@upd.ch
johann.kleinbub@gmail.com
volker.roth@unibas.ch
klaus.schmeck@upk.ch

[1] Psychiatric University Hospital, Basel, Switzerland
[2] University of Basel, Faculty of Psychology
[3] University of Kentucky, Department of Psychology
[4] University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland
[5] Section for Experimental Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre for Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany
[6]Section for Translational Psychobiology in Child and Adolescent Psychiatry, Department of Child and Adolescent Psychiatry, Centre for Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany
[7] Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy.
[8] University of Basel, Department of Mathematics and Informatics

**Corresponding Author:**
Martin Steppan Dr.
mhsteppan@gmail.com
Wilhelm Klein Strasse 27
Basel, Switzerland

**Conflicts of interest:**
The authors have no conflicts of interest to disclose

**Ethical review committee statement:** This study is part of the multi-centre study 'Evaluation of Adolescent Identity Treatment' registered at clinicaltrials.gov (NCT02518906). Ethical approval was obtained from the local ethics committee.

**Location:** All psychotherapies, video recordings and analyses were done at the Psychiatric University Hospital in Basel, Switzerland.

**Word Count:** 5,110 (including everything)

**Abstract**

Background: New advances in the field of machine learning make it possible to track facial emotional expression with high resolution, including micro-expressions. These advances have promising applications for psychotherapy research, since manual coding (e.g. the Facial Action Coding System [FACS]), is time-consuming. Purpose: We tested whether this technology can reliably identify in-session emotional expressions in a naturalistic treatment setting. Method: We applied a Convolutionary Neural Network (CNN) for emotion recognition to video material from 389 psychotherapy sessions of 23 patients with borderline personality pathology. We validated the findings with human ratings according to the Clients Emotional Arousal Scale (CEAS) and explored their associations with treatment outcomes. Results: Overall, machine learning ratings showed significant agreement with human CEAS ratings. Machine learning emotion recognition, particularly the display of positive emotions (smiling and happiness), showed medium effect size on improvement through treatment (d=0.3). Conclusions: Machine learning is a highly promising resource for process research in psychotherapy. The results highlight the differential associations of displayed positive and negative feelings with treatment outcomes. Machine learning emotion recognition may be used for the early identification of drop-out risks and clinically relevant processes in psychotherapy. Similar to research on psychiatric genetics, we suggest the sharing of video analysis summary data, to further investigate the potential of this method in psychiatry.

## Introduction

Facial expressions of emotion are evolved patterns, showing homology not only in humans and primates, but also in more distant species [1]. Despite this homology, the assessment of specific emotions in psychiatric research and psychotherapy is complex and time-consuming. In psychotherapeutic settings, questionnaires have been developed to describe emotional states, although these questionnaires are usually completed after the emotional experience, potentially leading to recall bias [2]. To address this limitation, Ekman developed the Facial Action Coding System (FACS), which allows for a standardized and more precise quantification of facial expressions of emotions from images and videos [3]. FACS has been applied to psychotherapy research [4], but it is almost impossible to manually rate the millions of frames that comprise videos of a full course of treatment from one patient, much less larger sample sizes. Hence, in psychiatric research, more indirect measures (e.g. vocal tone [5], heart rate variability [6], skin conductance [7], hormonal status [8], or neuroimaging [9]) have primarily been used to identify individuals' emotional arousal.

New technologies in the field of machine learning make it possible to compare facial expressions with large libraries of prototypical emotions (e.g. happiness, disgust, surprise) in almost real-time [10]. This possibility offers a whole new area of research, in which displayed emotions can be tracked both within sessions and over the full course of treatment without requiring the same time investment from coders. The first commercial software products (e.g. Facereader) to apply machine learning emotion recognition (MLER) have shown substantial agreement with Ekman's FACS [11]. Researchers have conducted a proof of concept study demonstrating the utility of such

MLER software to assess facial expressions of emotion in psychotherapeutic interventions among patients with borderline personality disorder (BPD) [12]. However, this study was based on only 29 psychotherapeutic sessions from 12 patients. Due to the novelty of this technology, as opposed to verbal emotion expression [13], facial emotion expression is still largely understudied in psychiatric research.

Artificial Intelligence (AI) is an emerging area of research within psychiatry [14]. In recent years, AI approaches have been used for the early diagnosis of psychiatric and neurodegenerative disorders (e.g. dementia [15]). AI approaches are also being used in the treatment of psychiatric disorders to personalize psychiatric treatments and promote high performance medicine [17] or precision psychiatry [18]. A prototypical example of this is an AI monitor suggested by Jan et al. which can predict Beck Depression Inventory II (BDI-II) scores from vocal and visual expressions [19]. AI approaches demonstrate at least three clear benefits to psychotherapy process research. First, they offer comprehensive standardization. AI algorithms produce a standardized form of measurement that can incorporate more relevant inputs and dimensions than human raters can effectively process [20]. Second, AI approaches save time and boost resolution. Given that AI can accurately measure emotional change, it could be used to complement or replace highly laborious human ratings with higher resolution information of emotions displayed by patients and therapists (for instance traditional FACS coding requires the manual scoring of about 30 frames of video for each second of therapy). Third, AI approaches are relatively more reliable. Since applied machine learning algorithms are deterministic, the same input always leads to the same result, theoretically producing perfect test-retest reliability.

To enhance the practical applicability of this research, researchers must demonstrate the external and criterion validity of AI (i.e. its utility for naturalistic treatment data). Thus, the aim of this study is to test (a) how MLER applied to videos of psychotherapy sessions compares to "gold standard" expert ratings of emotional arousal in patients; and (b) whether MLER shows clinically relevant associations with treatment outcomes. Because our lab has one of the largest video libraries of psychotherapeutic interventions on adolescent patients in Switzerland, we can provide more information on the quality of MLER as a potential tool in psychiatric and psychotherapeutic research.

## Methods

### *Ethical approval*

This study is part of the multi-centre study 'Evaluation of Adolescent Identity Treatment' that has been registered at clinicaltrials.gov (NCT02518906) [21, 22]. The current analyses are based on the entire available data collected at one participating center (Psychiatric University Hospitals, Basel). Ethical approval was obtained from the local ethics committee. All adolescents, their parents and the therapists provided written informed consent for participation.

### *Sample*

The sample consists of 23 adolescent patients with borderline personality pathology, defined as (1) meeting DSM-IV criteria for BPD using the Structured Clinical Interview for DSM-IV Axis II Personality Disorders; SCID-II [23] and (2) presenting with identity diffusion according to the Assessment of Identity Development in Adolescence (AIDA; total *T* score >60) [24, 25]. The mean age of the patients was 16.2 years (*SD* = 1.6), 21 (91.3%) out of 23 patients were female. The original material consisted of footage of 423 therapy sessions (duration approx. 50 min each). Thirty-four videos were excluded mainly due to errors in face detection (e.g. the algorithm consistently detected a background object as a face) leading to a final sample of 389 analyzed psychotherapy sessions.

### *Video processing and machine learning*

Figure 1 illustrates the video processing design. OpenCV was used to detect faces [26]. Since faces were relatively stationary throughout the therapy sessions, face detection was optimized by

dynamically defining an area of high likelihood of facial presence. If no face was detected for longer than one second, the region of interest was extended to ensure the accuracy of face detection. To validate this method, we saved an image of the selected region of interest every 100 frames for manual control of the procedure. To detect emotions, we implemented an open source model (https://bit.ly/2ERl3TO) based on a Convolutional Neural Network (CNN). This network was trained using the ”FER-2013 dataset” (https://bit.ly/3gQLm9T), a collection of 35,685 facial images designed to develop machine learning algorithms for facial emotion recognition[27]. Similar to previously used algorithms (e.g. the software FaceReader [12]), probabilities of the occurrence of six basic emotions (happiness, surprise, anger, disgust, sadness, fear), plus a percentage of neutral expression were extracted. Because we analysed more than one month of total video material, and to reduce computing time, calculations were performed at sciCORE (http://scicore.unibas.ch/), a high performance cluster computer and scientific computing center at the University of Basel.

*****

Figure 1 about here

*****

*Client Emotional Arousal Scale*

Three observers were trained to code videos using the Client Emotional Arousal Scale (CEAS). For each 1 min interval, raters code (1) the intensity of the patient’s arousal on a 1 (*low/no arousal*) to 7 (*full arousal*) ordinal Likert scale and (2) identify the predominant discrete emotion

(e.g. joy, sadness, fear, anger, disgust, love). To note, a discrete emotion was not identified for intervals rated low in emotional arousal (< 3). Raters first completed a training phase, in which 7 sessions were independently rated followed by a consensus rating. A total of 232 sessions were rated (on average 9.8 sessions per patient). Altogether 11,960 minute-intervals (i.e. 186.5 hours) of video material were rated manually. Inter-rater reliability of CEAS has been reported from 0.75 to 0.81 in previous studies [28 29].

*External criteria for reliability and validity*

To verify the validity of the MLER ratings, we validated the MLER ratings using the human-rated CEAS scores [30 31]. To test for clinical validity (i.e. the importance of MLER for therapeutic outcome) we regressed the MLER ratings on pre- and post-treatment measures of several psychological questionnaires. These included five subscales of the Youth Outcome Questionnaire (Y-OQ) [32]: Intrapersonal Distress, Somatic, Interpersonal Relationships, Social Problems, Behavioural Dysfunction; four subscales of the Level of Personality Functioning Questionnaire (LoPFQ 12-18)[33]: Self Direction, Empathy, Identity, and Intimacy; as well as the Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD) [34]. Six individuals dropped out of treatment, leaving 17 patients with treatment outcome data.

*Statistical analysis*

We first tested the degree of agreement between MLER and CEAS human ratings on unspecified emotional arousal using Pearson's correlations between CEAS ratings and the probability of a neutral facial expression according to MLER ("1-neutral" = nonspecific arousal). We controlled

for multilevel effects by including a random intercept to allow for different baseline levels in arousal for different patient-therapist dyads, using the R package *lme4* [35]. To provide information regarding objectivity of MLER, we report the correlation of MLER nonspecific arousal ("1-neutral") for three different human raters. Fisher's *z* test was performed to test whether individual rater correlations differed from the overall trend. We then showcase the temporal agreement of nonspecific CEAS-arousal ratings and MLER nonspecific arousal over the course of one session as an exemplar case.

For more specific human ratings using CEAS, we applied logistic regressions to determine how well MLER predicted human classifications of emotions. In addition to nonspecific CEAS ratings, coders rated five specific emotions: sadness, anger, joy, fear, and love (not enough information for disgust < 10 minutes rated video material). We fit logistic regressions for each emotion and each patient-therapist dyad separately, where the presence of an emotion was used as the dependent binary variable, and seven MLER scores were used as predictors. Sensitivity and specificity were calculated by adding up individual classification tables of each logistic regression model.

Regarding therapy outcome, we calculated a mean rank of every individual on all subscales of the questionnaires (YOQ, LoPF, ZAN-BPD; scored so higher scores indicate greater severity) before and after treatment. Change on these measures was calculated as the difference between these mean ranks, Δ. Using median-split we transformed this continuous delta also into two groups of "good outcome" (≥ Median), and "poor outcome" (< Median). To illustrate the association of MLER with therapy outcome, we plotted MLER against each minute (1-50) in all therapy sessions, and calculated independent samples *t*-tests, as well as Cohen's *d* for each comparison between the good outcome, poor outcome, and dropout group. Furthermore, we

applied a principal component analysis (PCA) to all seven MLER categories. The first two principal components (eigenvalues >1), were plotted for the different outcome groups (good, poor, dropout) to illustrate the relative localization of these groups within the space of positive and negative affectivity. Finally, we report a correlation matrix of MLER categories with binary and continuous therapy outcome, using MLER aggregated at the highest level (i.e. the level of patient-therapist dyads).

Except for the between-patient correlations, we analyzed relations involving MLER by averaging MLER scores across blocks of minutes. This choice allowed us to directly compare MLER ratings with human CEAS ratings which were made across minute intervals. All analyses and visualization were done in R Studio using R (Version 3.5.3) [36]. We used the R package ggplot2 for all visualizations [37].

## Results

*Agreement between human raters and MLER*

Figure 2 illustrates the agreement of human raters and machine learning emotion recognition. Video sequences, which were rated as relatively highly emotionally arousing by humans (CEAS>3), were also classified as less neutral by MLER. Figure 2A shows an example of one session with comparatively high agreement between human ratings and MLER ($r$ = .50). Figure 2B illustrates the overall correlation between MLER and human CEAS ratings. The overall correlation (irrespective of rater) was $r$ = .08, $p$ < .01. Figure 2B also shows the agreement of MLER with human ratings, broken down by three human raters who evaluated the video material. For all three human raters, the association was positive and significant (see also Supplementary Figure S2). When applying a random effects model, controlling for baseline differences between raters and patients, the effect remained highly significant ($p$s < .01). When comparing each human rater to the overall agreement between CEAS and MLER (Fisher's $z$), all three raters showed essentially similar slopes (see Fig. 2B).

*****

Figure 2 about here

*****

Table 1 illustrates the relative utility of each of the seven MLER emotions for classifying specific emotions as labelled by human raters. Whereas nonspecific emotional arousal showed only moderate correct classification (72.3%), specific emotional states like sadness, anger, joy, and love each were correctly classified on > 90% of occasions. A large part of this high accuracy

is due to high specificity (detection of true negatives), which again exceeded 90% for all specific emotions. Only "love", which was labelled in 86 out of 10,736 rated minute intervals, also showed a perfect sensitivity (100%) (i.e. also a correct classification of true positives).

*****

Table 1 about here

*****

*Association between MLER and the outcome of psychotherapy*

Figure 3 (A-G) illustrates the display of emotions within the course of a given session, stratified by three outcome groups (good outcome, poor outcome, dropout). All three groups show relatively distinct patterns, as well as some similar features regarding the display of basic emotions. For example, at the beginning of each session, all three groups show a relatively high level of happiness, which decreases after about 5 minutes. Conversely, sadness and fear are relatively lower at the beginning of each session, and only reach their average level (horizontal lines) after 15-20 minutes.

We also found distinct patterns of emotional expressions between the good outcome (N=8), poor outcome (N=9), and one dropout groups (N=6). For instance, all MLER emotional categories except surprise, demonstrated significant between-group differences ($d$s > .10). The most pronounced effect can be seen for happiness, which is significantly higher in patients with a good therapy outcome, compared to those with a poor outcome, $d = 0.30$, $p < .01$, and those who dropped out ($d = 0.32$, $p < .01$. In contrast, the poor outcome group demonstrated consistently and significantly more arousing sadness, anger, disgust and fear compared to the good outcome

group. This corresponds with a Principal Components Analysis (Figure 3H). Based on component loadings (see Appendix Table S2), the first two principal components can be described as positive affectivity (positive loadings on happiness and nonspecific arousal) and negative affectivity (positive loadings on anger, sadness, fear and nonspecific arousal). When plotting the factor scores of every minute measured using MLER, the three outcome groups show relatively distinct, albeitoverlapping areas of higher presence. Whereas the good outcome group (green dots) is more often located in the bottom right quadrant (positive affectivity), the poor outcome group shows a higher density in the upper two quadrants (negative affectivity). Those in the dropout group also demonstrated scores in an area of higher density in the lower left quadrant, which can be described as neither positive nor negative affectivity. This is also in line with a generally reduced overall nonspecific arousal of the dropout group compared to both other groups (vs. good outcome: $d = 0.18$, $p < .01$; vs. poor outcome: $d = 0.17$, $p < .01$).

*****

Figure 3 about here

*****

*****

Table 2 about here

*****

Table 2 shows a conservative estimate of potential associations of MLER with pre- and post-treatment symptomatology, dropout, as well as the continuous improvement over therapy, Δ. Aggregated at the highest level (individuals), the display of happiness was significantly (negatively) correlated, $r = -.53$, $p < .01$, with post-treatment symptomatology, but not pre-treatment symptomatology. Similarly, individual improvement, Δ, was significantly and positively correlated, $r = .49$, $p < .02$, with the display of happiness in treatment. For several other cells, numerically relevant, but not significant association are reported here, suggesting a general tendency for negative affectivity to be associated with less improvement (negative correlations for sadness, anger, disgust and fear), as well as with lower nonspecific arousal (negative correlation with neutral).

**Discussion**

In a naturalistic setting of 389 psychotherapy sessions, we tested the utility of machine learning emotion recognition for psychotherapy process research. We found substantial evidence corroborating the standardization of the technology, as well as strong evidence for its external and ecological validity, given significant and meaningful associations with outcome and dropout from psychotherapy.

Several limitations of this approach are worth considering. First, the applied machine learning algorithms are a patchwork of open source algorithms. More specialized software products may exist, for example in the area of advanced robotics [38]. Hence, the applied MLER may still underestimate the true potential of this methodology. Second, the machine learning algorithms we applied were not tailored to individual patients. Stronger associations can be expected if the algorithms were trained specifically to every individual. However, we did not use this approach to avoid "overfitting" and to increase the generalizability of the results. Third, the sample we used is relatively specific (i.e. adolescents with borderline personality pathology). Hence, results might be different in other samples. The comparison of emotional expression as measured by MLER, and its association with outcomes in different patient groups and different treatments appears to be a formidable direction for future research. Fourth, a fundamental statistical limitation of this analysis is that data is nested within 23 individuals for which therapy outcome and dropout information were available. However, we only found robust associations with post-treatment symptomatology and change over time, but not with pre-treatment symptomatology. Hence, the results suggest that the display of positive emotions is more a function of the treatment and not of the pre-existing personality disposition of the individuals. Fifth, despite the fast processing time of data relative to human coders, the MLER approach used in this study

must still be considered overly computationally intensive. Extraction of the data would not have been possible in such a short time without the mentioned scientific computing center.

When comparing MLER to a "gold standard" of human ratings (CEAS) of patients' facial expressions of emotions in psychotherapy [29], we found stable correlations with specific emotions, as well as with nonspecific emotional arousal. Together with previous studies that have shown agreement between MLER and human ratings of emotions [11,39], these results provide evidence that MLER indeed demonstrates adequate criterion validity (i.e. it measures what it is intended to measure) [40]. When comparing the agreement of MLER with human ratings for different human raters, we found that the slopes were consistently positive and individual raters' slopes do not differ significantly from the overall association. These findings indicate that MLER demonstrates observer invariance (i.e. it is largely independent of the observer with which it is compared). However, it is noteworthy that the correlations between MLER and human ratings are numerically low. This may be due to the fact that human ratings were done for intervals of several minutes at a time. Within this period, MLER can have a much higher resolution, given a maximum of 1,800 frames (up to 30 frames / second). This mismatch in the resolution of MLER and the criterion (CEAS), may explain why these correlations are small-sized while also being statistically significant.

There is little research to compare our findings to, as this represents the first study to apply MLER to all treatment sessions. The results are in line with earlier findings applying the FACS to treatment sessions, highlighting the importance of smiling and positive feelings in psychotherapy [4]. By contrast, Arango et al. (2019) found that expressions of happiness and fear

became more intense and sadness less intense over the course of a session in 29 therapy sessions of 12 patients with BPD in Mexico [12]. Despite testing the same population of patients with BPD, these discrepancies may indicate cultural differences in emotion expression [41].

To test the ecological validity of emotional expressions, we explored the associations between MLER and treatment outcomes, We found small- to medium-sized associations between MLER-rated expressions of all emotions except fear and change in outcomes across treatment. For example, our results to some extent align with previous studies and meta-analyses, showing that emotion suppression has a negative impact on the outcome of psychotherapies [13 42]. Although we did not find associations with pre-treatment symptomatology, there were stable group differences regarding post-treatment symptomatology, dropout, and individual improvement ("delta"). These results suggest that emotional expressions may also have prognostic value for treatment outcomes. This does not necessarily imply that specific emotions are causal agents. However, intensities of emotional expressions may be indicators of causal agents, such as discomfort in treatment or the stability of a patient-therapist interactions. For instance, results from a previous meta-analysis showed that the personal relationship between patient and therapist played a crucial role in treatment outcomes [43], which may be reflected by the display of more positive emotions in this study.

This pilot study contributes to the development and application of machine learning emotion recognition for psychiatric research. We provided further evidence that MLER met several fundamental quality criteria (reliability, objectivity, validity), suggesting it could complement or even outperform traditional emotion recognition approaches based on physiological measures (e.g. skin conductance, heart rate variability). MLER may have clinical applications in the

identification of highly salient emotional sequences. As computational requirements are reduced through algorithmic refinement, MLER may be used in idiographic studies, in the supervision of mental health experts, and potentially as a clinical tool, highlighting important emotional phases which may go unnoticed, and offering an enhanced experience for patients in psychotherapy that could contribute to improved outcomes. Based on the encouraging results of this study, we suggest the pooling of international efforts. Similar to research in psychiatric genetics, the pooling of anonymized summary data on MLER, would increase the statistical power and would help to further elucidate the potential of this method on a wider range of disorders and age groups. We invite research groups who have collected video material of psychotherapeutic treatment, to get in touch.

**Acknowledgments**

Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at University of Basel.

**Financial Disclosure**

No financial conflicts to disclose.

**References**


1. Waller BM, Julle-Daniere E, Micheletta J. Measuring the evolution of facial 'expression' using multi-species FACS. *Neurosci & Biobehav Rev*. 2020.

2. Schiepek G, Stöger-Schmidinger B, Kronberger H, et al. The Therapy Process Questionnaire-Factor analysis and psychometric properties of a multidimensional self-rating scale for high-frequency monitoring of psychotherapeutic processes. *Clin Psychol & Psychother*. 2019;26:586-602.

3. Ekman P, Rosenberg EL. *What the Face RevealsBasic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press; 2005. http://dx.doi.org/10.1093/acprof:oso/9780195179644.001.0001. Accessed 2005.

4. Bänninger-Huber E. Prototypical Affective Microsequences in Psychotherapeutic Interaction. *Psychother Res*. 1992;2(4):291-306. doi:10.1080/10503309212331333044

5. Soma CS, Baucom BR, Xiao B, et al. Coregulation of therapist and client emotion during psychotherapy. *Psychother Res*. 2020;30:591-603.

6. Dikecligil GN, Mujica-Parodi LR. Ambulatory and challenge-associated heart rate variability measures predict cardiac responses to real-world acute emotional stress. *Biol Psychiatry*. 2010;67:1185-1190.

7. Kleinbub JR. State of the art of interpersonal physiology in psychotherapy: a systematic review. *Front Psychol*. 2017;8:2053.

8. Schumacher S, Niemeyer H, Engel S, Cwik JC, Knaevelsrud C. Psychotherapeutic treatment and HPA axis regulation in posttraumatic stress disorder: A systematic review and meta-analysis. *Psychoneuroendocrinology*. 2018;98:186-201.

9. Oathes DJ, Ray WJ, Yamasaki AS, et al. Worry, generalized anxiety disorder, and emotion: Evidence from the EEG gamma band. *Biol Psychol*. 2008;79:165-170.

10. Jain N, Kumar S, Kumar A, Shamsolmoali P, Zareapoor M. Hybrid deep neural networks for face emotion recognition. *Pattern Recognit Lett*. 2018;115:101-106.

11. Skiendziel T, Rösch AG, Schultheiss OC. Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. *PloS one*. 2019;14:0223905.

12. Arango I, Miranda E, Sánchez Ferrer JC, et al. Changes in facial emotion expression during a psychotherapeutic intervention for patients with borderline personality disorder. *J Psychiatr Res*. 2019;114:126-132. doi:10.1016/j.jpsychires.2019.04.026

13. Peluso PR, Freund RR. Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. *Psychotherapy*. 2018;55(4):461-472. doi:10.1037/pst0000165

14. Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. *Mol Psychiatry*. 2019;24:1583-1598.

15. Egger K, Rijntjes M. Big data and artificial intelligence for diagnostic decision support in atypical dementia. *Der Nervenarzt*. 2018;89:875-884.

16. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Internet Res*. 2019;21:13216.

17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56.

18. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry: Cogn Neurosci Neuroimaging*. 2018;3:223-230.

19. Jan A, Meng H, Gaus YFBA, Zhang F. Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. *IEEE Trans Cogn Dev Syst*. 2017;10:668-680.

20. Dawes TR, Eden-Green B, Rosten C, et al. Objectively measuring pain using facial expression: is the technology finally ready? *Pain Manag*. 2018;8:105-113.

21. Schmeck K, Pick OG, Milidou M, Schenk N, Schlüter-Müller S, Zimmermann R. Früherkennung von Persönlichkeitsstörungen. *PTT-Persönlichkeitsstörungen: Theor und Ther*. 2018;22:179-185.

22. Zimmermann R, Fürer L, Schenk N, et al. Silence in the psychotherapy of adolescents with borderline personality pathology. *Personal Disord Theory, Res Treat*. 2020. doi:10.1037/per0000402

23. First MB, Spitzer RL, Gibbon M, et al. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part II: Multi-Site Test-Retest Reliability Study. *J Personal Disord*. 1995;9(2):92-104. doi:10.1521/pedi.1995.9.2.92

24. Goth K, Foelsch P, Schlüter-Müller S, et al. Assessment of identity development and identity diffusion in adolescence - Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child Adolesc Psychiatry Ment Health*. 2012;6(1):27. doi:10.1186/1753-2000-6-27

25. Lind M, Vanwoerden S, Penner F, Sharp C. Inpatient adolescents with borderline personality disorder features: Identity diffusion and narrative incoherence. *Personal Disord Theory, Res Treat*. 2019;10(4):389-393. doi:10.1037/per0000338

26. Zelinsky A. Learning OpenCV---Computer Vision with the OpenCV Library (Bradski, G.R.

et al.; 2008)[On the Shelf]. *IEEE Robot & Autom Mag*. 2009;16(3):100-100. doi:10.1109/mra.2009.933612

27. Goodfellow IJ, Erhan D, Luc Carrier P, et al. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*. 2015;64:59-63. doi:10.1016/j.neunet.2014.09.005

28. Missirlian TM, Toukmanian SG, Warwar SH, Greenberg LS. Emotional arousal, client perceptual processing, and the working alliance in experiential psychotherapy for depression. *J Consult Clin Psychol*. 2005;73:861.

29. Greenberg LS, Auszra L, Herrmann IR. The relationship among emotional productivity, emotional arousal and outcome in experiential therapy of depression. *Psychother Res*. 2007;17:482-493.

30. Rosner R. *The Relationship between Emotional Expression, Treatment and Outcome in Psychotherapy: An Empirical Study*. Vol 565. Peter Lang Pub Incorporated; 1996.

31. Machado P. Clients emotional arousal in therapy: Development of a rating scale. *Unpubl Manuscr*. 1992.

32. Burlingame GM, Wells MG, Lambert MJ, Cox JC. Youth outcome questionnaire (Y-OQ). *Use Psychol Test Treat Plan Outcome Assessment, ed*. 2004;3:235-274.

33. Goth K, Birkhölzer M, Schmeck K. Assessment of Personality Functioning in Adolescents With the LoPF–Q 12–18 Self-Report Questionnaire. *J Personal Assess*. 2018;100(6):680-690. doi:10.1080/00223891.2018.1489258

34. Zanarini MC. Zanarini Rating Scale For Borderline Personality Disorder (ZAN-BPD): A Continuous Measure of DSM-IV Borderline Psychopathology. *J Personal Disord*. 2003;17(3):233-242. doi:10.1521/pedi.17.3.233.22147

35. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv. org. *Prepr doi*. 2014;10.

36. Team Rs. RStudio: integrated development for R. *RStudio, Inc, Boston, MA URL http: //www rstudio com*. 2015;42:14.

37. Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*. 2011;67(2):678-679. doi:10.1111/j.1541-0420.2011.01616.x

38. Ruiz-Garcia A, Elshaw M, Altahhan A, Palade V. A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots. *Neural Comput Appl*. 2018;29:359-373.

39. Haines NA, Southward MW, Cheavens JS, Beauchaine T, Ahn WY. Using computer vision and machine-learning to automate facial coding of positive and negative affect intensity. *PLoS ONE*. 2019 Feb 5;14(2):e0211735. doi: 10.1371/journal.pone.0211735.

40. Bornovalova MA, Choate AM, Fatimah H, Petersen KJ, Wiernik BM. Appropriate use of bifactor analysis in psychopathology research: appreciating benefits and limitations. *Biol Psychiatry*. 2020.

41. Geiger M, Wilhelm O. Computerized Facial Emotion Expression Recognition. *Digit Phenotyping Mob Sens*. 2019:31-44.

42. Scherer A, Boecker M, Pawelzik M, Gauggel S, Forkmann T. Emotion suppression, not reappraisal, predicts psychotherapy outcome. *Psychother Res*. 2017;27:143-153.

43. Gelso CJ, Kivlighan Jr DM, Markin RD. The real relationship and its role in psychotherapy outcome: A meta-analysis. *Psychotherapy*. 2018;55:434.

**Legends**

*Figure 1*. Design for image processing. 1= to reduce error, a region of interest (ROI) can be defined where a face is detected with high likelihood which is expanded if face detection fails; 2=Faces are detected and transformed into 48x48 pixels grayscale pictures; 3= The convolutional neural net had been trained using the FER-2013 dataset.

*Figure 2*. Agreement between human ratings (CEAS) and MLER nonspecific emotional arousal ("1: no arousal"). Figure 2A: Example plot of one session with high agreement. Figure 2B: Overall agreement between CEAS and MLER for all cases. Colors indicate different human raters).

*Figure 3*. A-G: MLER emotional expression within all analyzed therapy sessions (per minute) by therapy outcome group (green = good, blue = poor, red = dropout; vertical lines indicate standard errors; z scores over all values). Test statistics are shown for each group comparison. H: Scatter plot showing the results of a Principal Component Analysis on all minute intervals where MLER scores were available.

*Table 1*. Logistic regressions to classify human ratings using MLER

*Table 2*. Correlation matrix of MLER with treatment outcome measures

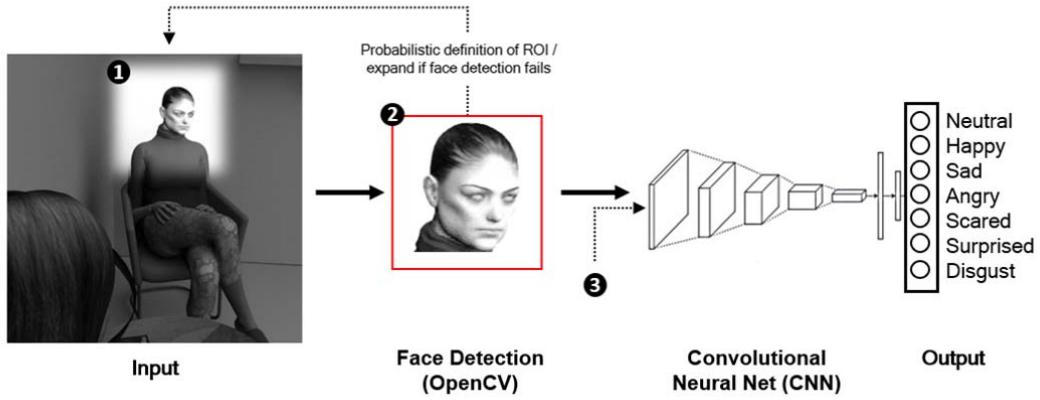


*Figure 1*. Design for image processing. 1= to reduce error, a region of interest (ROI) can be defined where a face is detected with high likelihood which is expanded if face detection fails; 2=Faces are detected and transformed into 48x48 pixels grayscale pictures; 3= The convolutional neural net had been trained using the FER-2013 dataset.

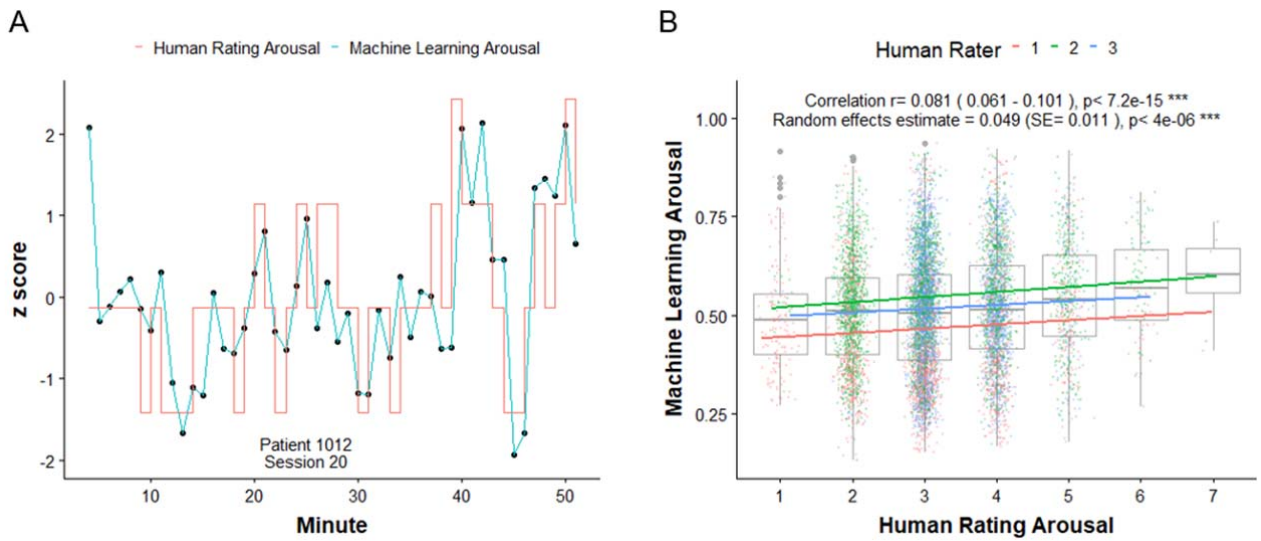


*Figure 2*. Agreement between human ratings (CEAS) and MLER nonspecific emotional arousal ("1: no arousal"). Figure 2A: Example plot of one session with high agreement. Figure 2B: Overall agreement between CEAS and MLER for all cases. Colors indicate different human raters).

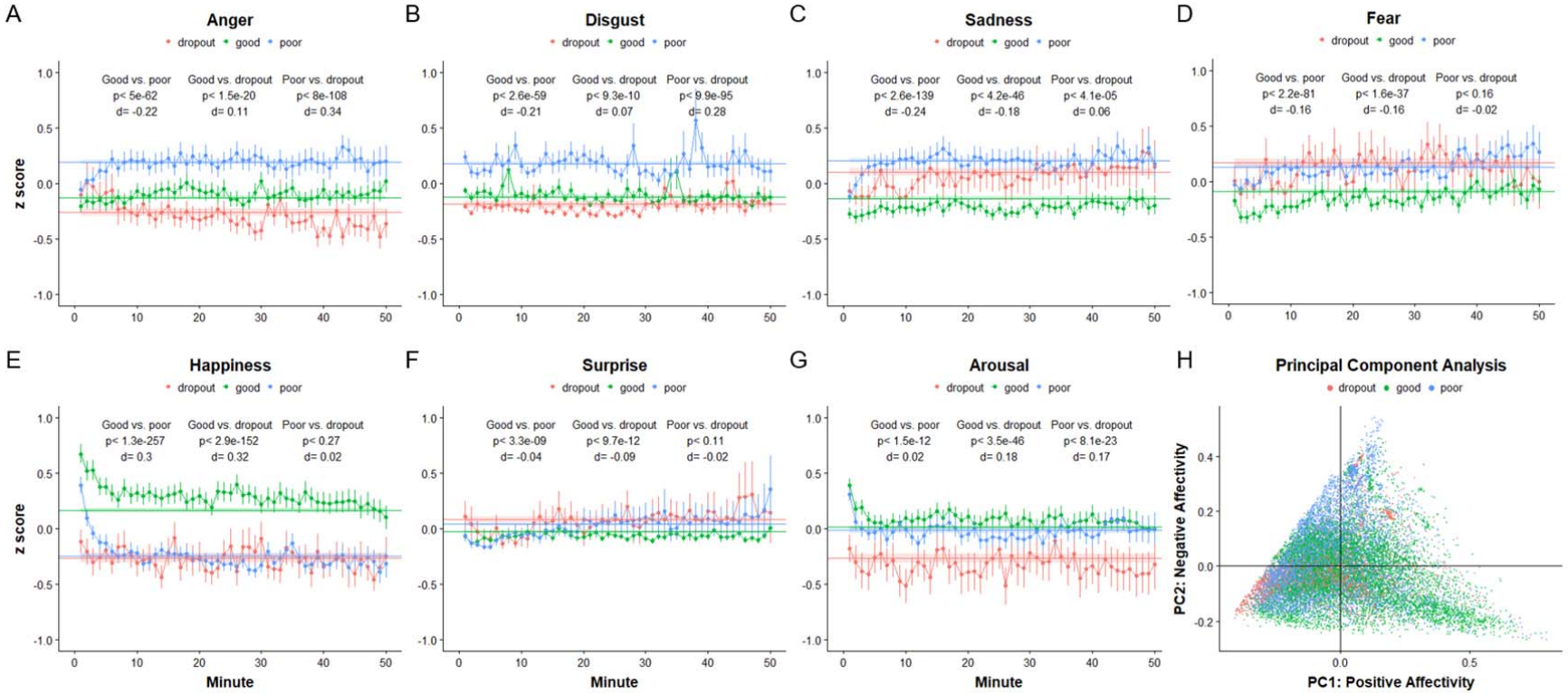


*Figure 3*. A-G: MLER emotional expression within all analyzed therapy sessions (per minute) by therapy outcome group (green = good, blue = poor, red = dropout; vertical lines indicate standard errors; z scores over all values). Test statistics are shown for each group comparison. H: Scatter plot showing the results of a Principal Component Analysis on all minute intervals where MLER scores were available.

**Table 1.** Binary-logistic regressions to classify human ratings using MLER

| Human Rating | Correct classification | Sensitivity | Specificity | Minutes |
|---|---|---|---|---|
| Neutral (CEAS<4) | 72.3% | 74.6% | 63.5% | 6121 |
| Sadness | 90.3% | 66.2% | 90.9% | 891 |
| Anger | 90.2% | 65.0% | 90.4% | 888 |
| Joy | 95.8% | 68.0% | 95.9% | 384 |
| Fear | 91.4% | 55.6% | 92.1% | 789 |
| Love | 98.3% | 100.0% | 98.3% | 86 |

*Note.* Regression models fitted for each individual separately using MLER emotions as predictors (anger, sadness, happiness, disgust, fear, surprise, neutral). Classification summed up over all individuals.

**Table 2.** Correlation matrix MLER and the outcome of psychotherapy

| Emotion | Pre | Post | Delta | Dropout |
|---|---|---|---|---|
| Angry | -0.12 | 0.32 | 0.29 | -0.27 |
| Disgust | -0.33 | -0.05 | -0.23 | -0.33 |
| Scared | 0.26 | 0.26 | -0.05 | 0.15 |
| Happy | 0.06 | **-0.53*** | **0.49*** | -0.14 |
| Sad | 0.00 | 0.33 | -0.42 | 0.02 |
| Surprised | 0.31 | 0.37 | -0.13 | 0.11 |
| Unspecific arousal | 0.07 | -0.15 | 0.15 | -0.21 |

*Note.* Pearson correlations. MLER aggregated on the level of individuals; * p<.05; Delta= Pre – Post symptomatology rank; Unspecific arousal = “1-neutral”

**S4: Alliance Ruptures and Resolutions in Personality Disorders**

Qualitative research in the events paradigm on the ongoing negotiation of the therapeutic relationship has introduced several rating systems, including the rupture-resolution rating system, already presented in the sections ahead (Eubanks et al., 2019). A distinct doctorate programme has been dedicated to the study of ruptures and resolutions in the here used data (Schenk et al., 2019, 2020). In accordance with proposed deficits in self- and interpersonal functioning and being a defensive by-product of the here relevant working process of identity integration, it is discussed that especially patients with BPD are prone to have difficulties in the establishment of a benign therapeutic relationship (Linehan et al., 2000). In part, strong countertransference reactions (anger, confusion, anxiety) challenge the building and maintenance of a benign therapeutic relationship (McMain et al., 2015; see also secondary emotion: Greenberg & Safran, 1989). It is hypothesized that patients with BPD are more prone to show ruptures. Further, according to the relational approach, the resolution of ruptures in patients with BPD – and personality disorders in general – not only ensures a healthy relationship, but rather is hypothesized to be a change mechanism in itself (Safran, 1993). Applying expressive resolution strategies (Safran & Muran, 1996), ruptures offer a window for exploring the patients' core relational themes as they unfold in the therapeutic relationship (Muran & Safran, 2017). However, literature on the special role of rupture repair in BPD compared to other psychiatric disorders is scarce. Due to the working group's expertise, we got invited to write a review article on the effects of rupture repair in personality disorders in general.

# Alliance Ruptures and Resolutions in Personality Disorders

Nathalie Schenk[1*], Lukas Fürer[1*], Ronan Zimmermann[1,2], Martin Steppan[1], Klaus Schmeck[1]

[1]Child and Adolescent Psychiatric Research Department, Psychiatric University Hospitals of the University of Basel, Basel, Switzerland

[2]Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of Basel

* Equal contribution

**Abstract:**

***Purpose of review:*** This review provides an overview of the state of research on alliance ruptures and resolutions in the treatment of personality disorders (PD). We discuss frequently used instruments to measure alliance ruptures and resolutions. We discuss the effectiveness of rupture resolution processes and highlight possible avenues for research to explore. Innovative assessments with the potential to reveal the link of ruptures and resolutions and mechanisms of psychotherapeutic change are discussed.

***Current findings*:** The assessment of alliance rupture and resolutions is heterogeneous. Instruments vary largely with respect to a direct or indirect assessment, the time resolution of assessment (integral therapy, phase, session, event), session sampling strategy and perspectives (patient, therapist, observer). The heterogeneity in the instruments and study designs impedes comparability and interpretation of the findings. Results support the hypothesis that ruptures are more frequent in PD. Results also point towards beneficial effects of rupture resolution patterns, early alliance quality and resolution complexity. Few studies control findings for pretreatment factors.

***Summary*:** Evidence points to the direction that rupture resolution processes can be considered a general principle of change in the treatment of PD. The concept of alliance ruptures and resolutions provides a useful tool for the management of the therapeutic alliance and its moments of deteriorations throughout the treatment course. Dimensional pretreatment personality functioning is considered a key variable in future studies to highlight what works for whom.

Keywords: Rupture Resolution, Personality Disorder, Alliance

# 1 Introduction

It is argued that relationship factors are not only a byproduct of the psychotherapeutic encounter but a vehicle of change per se (Task Force Division 29: [1, 2]). Building upon a relational approach [••3], alliance research concerned with personality disorders (PD) investigates the assumption that the process of rupture resolution brings about change [4, 5]. The NICE guidelines recommend the building and management of the therapeutic alliance as a general principle of change in the treatment of PD [6]. However, patients with PD have general difficulties in forming interpersonal relationships, including the therapeutic relationship, due to their impairments in self and interpersonal functioning [7].

Safran & Muran [8] conceptualize the therapeutic alliance as a dynamic and relational entity that is continuously built and negotiated by patient and therapist. In this negotiation process alliance ruptures with minor or major significance emerge inevitably. Alliance ruptures are described as momentary deteriorations in the therapeutic alliance resulting from a lack of collaboration regarding the goals and tasks of the treatment or resulting from strains in the emotional bond [•9]. Withdrawal and confrontation ruptures can be differentiated by their movement dynamics (against therapist or therapy work versus avoiding therapist or therapy work). The rupture resolution is the process of recognizing, exploring and repairing alliance ruptures with the aim to reestablish the collaboration and to explore the patient's underlying needs and intrapsychic conflicts. The resolution process that is initiated by the therapist has the potential to enable a corrective experience that is held to be therapeutic in itself [4]. Safran & Muran [5] suggest that alliance ruptures disclose a window into core interpersonal schemes. Insofar, the therapeutic relationship enables an interpersonal learning field in which representations of the self and others, interpersonal schemes and the needs for agency versus communion can be probed and adapted [••3]. Resolution strategies can be immediate, focusing on the reestablishment of the collaboration, or expressive, exploring the patients' needs and intrapsychic conflicts underlying the alliance rupture [10].

Schenk et al. [•18] provide a detailed description of the occurrences of ruptures and resolutions in therapies with adolescents with borderline personality pathology (BPD). The most typical alliance ruptures were minimal response, denial, avoidant storytelling, patient defends self against therapist, complaints against therapist and rejecting therapist intervention. Therapists typically

used the resolution strategies to invite the patient to discuss thoughts and feelings about the therapist or some aspect of therapy, to validate the patient's defensive posture and to illustrate tasks or provide a rationale for treatment (Schenk et al. [•18]). The following transcript demonstrates a short rupture-resolution sequence of a therapy session with a 16-year-old adolescent with BPD:

T: How are you?
P: Not good.
T: Not good?
P: No. (Pause) => Minimal response rupture
T: what happened?
P: I don't know. (Pause) => Minimal response rupture
T: Is anything bothering you?
P: No. => Denial Rupture
T: How does it come you feel not good?
P: I don't know, it's like always. => Minimal response rupture
T: And what does it mean when you say that you are not doing well, any thoughts or feelings about that? => Resolution "therapist invites the patient to discuss thoughts or feelings"
P: I don't feel happy, just empty.
T: Mmmh. Do you already know that feeling from yourself?
P: I don't know. (Long pause) => Minimal response rupture
T: Hmm, I have the feeling we are stuck here right now. I have the feeling you could feel pressured by me asking you all these questions. Do you feel this way with me?
=> Resolution attempt "the therapist discloses his/her internal experience of the patient-therapist interaction"

## 2 Instruments of Alliance Ruptures and Resolutions

Alliance ruptures and resolutions can be measured with direct or indirect instruments, through different perspectives (patient, therapist observer), on different assessment levels (session, specific time windows, episodes, markers, speaking turns, utterances) and at different time points (successive therapy sessions, phases, specific sessions, between sessions, retrospective in interviews).

Table 1 presents frequently used direct and indirect instruments of alliance ruptures. The *direct* assessment is based on observer ratings or retrospective self-report questionnaires aiming directly at rupture resolution episodes. The *indirect* assessment means that the alliance is measured and inferences towards alliance ruptures and resolutions are drawn from sudden and significant

fluctuations in the global alliance. Eubanks-Carter et al. [11] provide an overview of the quantitative naturalistic methods for detecting alliance ruptures on the basis of global alliance measures.

The *direct* instruments with observer ratings (e.g. 3RS, CIS) enable a detailed investigation of the process of the alliance negotiation based on the event paradigm [12]. This fine-grained analysis is particularly of interest to understand underlying principles of change through rupture resolution processes. Using task analysis, the performance of rupture resolutions can be studied in detail as done by Aspland et al. [13], Bennett et al. [•14] and Daly et al. [15]. Furthermore, with descriptive studies the incidence and phenomenological quality of ruptures and resolutions can be examined in different patient samples [16-•18]. Limitations are that the coding is time and cost intensive and therefore often restricted to small sample sizes which impedes generalizability and inferences to the treatment outcome. The *indirect* instruments derived from self-report ratings of the alliance assess changes in the quality of the alliance on a global level. They estimate the impact of possible but not directly measured ruptures and likely measure major but not minor ruptures. Indirect measures allow the investigation of large patient samples on a macro level of investigation. They are feasible for the study of within-patient and between-patient effects of rupture resolution processes in relation to the treatment outcome. However, the indirect measures are less accurate and might be biased regarding construct validity. In comparison to observer ratings, self-report measures underestimate the amount of alliance ruptures [16, 19].

--- insert Table 1 here ---

# 3 Influence of PD traits on rupture resolution processes

Theoretical and clinical considerations about the interpersonal nature of different PDs allow assumptions towards the quality of ruptures encountered with these patients [•20, ••3]. For example, withdrawal ruptures may be more frequent in patients who are overly compliant, fearful and averse to interpersonal conflicts such as cluster C patients. Confrontation ruptures may be more frequent in patients with cluster A or B PD as they use more direct and overt strategies to disclose strains in the alliance like criticize or pressure the therapist. Unfortunately, there is no empirical evidence emphasizing these hypotheses.

Zilcha-Mano [••21] proposed to differentiate between trait-like (i.e. pretreatment tendencies of the patient to form relationships with others) and state-like components of alliance (i.e. changes in relationship functioning through the interaction with the therapist). It is hypothesized that patients with PD present with lower trait-like pretreatment interpersonal functioning that leads to problems in building and maintaining a stable working alliance. This could lead to more ruptures during the process and a lower (early) alliance quality compared to other clinical groups.

In the validation study of the 3RS Eubanks et al. [22] found ruptures in almost every session of a psychotherapy with patients with varying diagnoses. However, Schenk et al. [•18] used the 3RS to analyze the complete treatment of a sample of ten adolescents with BPD. In 72% of the sessions at least one rupture was observed.

Colli et al. [23] compared 15 PD patients (cluster B & C) with 15 patients suffering from other psychiatric disorders. In PD patients rupture markers were observed more frequently than in non-PD patients (tasks & goals marker: $F = 4.7$, $p = .031$, discouragement marker: $F = 4.0$, $p = .046$). Interestingly, they also found that therapists of PD patients tended to be more hostile, less clear and used more perseverations than non-PD therapists. However, the PD therapists also provided more supportive, explicative and expressive interventions targeting alliance ruptures than the non-PD therapists.

Tufekcioglu et al. [24] tested whether cluster C PD or personality traits impact the early alliance compared to non-PD patients. They found that PD patients rated a higher rupture intensity than non-PD patients ($F = 16.6$, $r = .43$, $p < .01$). Furthermore, they found that the pretreatment personality traits of high impulsivity, dysregulation and lability (but not diagnoses) were associated with higher patient-reported rupture intensity. Coutinho et al. [16] found that patients with PD experience a lower alliance quality (measured with the WAI) and more frequent withdrawal and confrontation ruptures (3RS) compared to patients with depression and anxiety disorders.

In conclusion, there is evidence that confrontation and withdrawal ruptures are more frequent in PD than other clinical samples. Differences between PD and non-PD patients seem to also surface when observing therapist behavior. The findings can be interpreted in the light of a trait-like alliance component of lower interpersonal functioning in PD patients, meaning that the quality of

the therapeutic relationship is endangered from the beginning on by the severe interpersonal problems that are at the core of PD symptomatology.

# 5 The Effects of Ruptures and Resolutions

Studies concerned with the correlation between alliance and outcome focus on different characteristics of the alliance, depending on the instruments used. We review the outcomes of meta-analyses, literature investigating early alliance, rupture resolution processes (e.g. significant fluctuations in WAI) and literature focusing on resolutions and their complexity. When trying to unveil the state-like components of alliance leading to change (curative interactions with therapists), one has to control for trait-like components (e.g. pretreatment personality functioning), setting the preconditions for a beneficial interpersonal relationship. Not all studies reviewed here do that.

Up to date there are two meta-analyses that report moderate effects of rupture resolution processes on the treatment outcome [•9, 25]. Safran et al. [25] summarized studies using indirect instruments of alliance ruptures. In this meta-analysis the number of rupture resolution episodes was positively related to treatment outcome ($r$ = .24, 95% *CI* [.09, .39], $p$ = .002, $k$ = 3, $n$ = 148). The meta-analysis of Eubanks et al. [•9] confirmed that rupture resolution processes are positively associated with the treatment outcome when using direct as well indirect measures ($r$ = .29, *CI* [.10, .47], $d$ = .62, $p$ = .003, $k$ = 11, $n$ = 1314). The effect was independent of type of treatment or type of instruments used. However, the timing of the assessment impacted the rupture resolution outcome relationship. An assessment in early sessions resulted in a lower association ($r$ =.13, $z$ = 1.90, $p$ = .06), whereas an assessment across the complete course resulted in a stronger rupture resolution outcome association ($r$ = .38, $z$ = 3.13, $p$ = .002). This is an encouraging result when hypothesizing that late sessions may better reflect state-like alliance components, whereas early sessions could be more influenced by the trait-like pretreatment conditions.

Considering early alliance, Strauss et al. [26] found that patients (obsessive-compulsive PD & avoidant PD) who showed higher pretreatment personality pathology presented with a lower early alliance ($r$ = -.40, $p$ < .05), and in turn, better early alliance predicted a higher number of sessions attended ($r$ = .38, $p$ < .05). Also, better early alliance predicted gains in all post treatment outcome

measures (personality pathology: $r = -.40$, $p < .05$, depression score: $r = -.49$, $p < .01$). Also, Muran et al. [27] found that lower early rupture intensity was associated with higher post-treatment interpersonal functioning (cluster C & NOS, patient rated: $r = -.35$, $p < .01$, therapist rated: $r = -.32$, $p < .01$). In a youth BPD sample, Gersh et al. [28] found that more early ruptures were associated with poorer outcomes in social functioning ($r = .32$, $p < .05$).

There is evidence for rupture resolution patterns (significant fluctuations or dips in alliance implying ruptures, returning to baseline and above, implying resolution). Schenk et al. [•18] assessed ruptures with the 3RS on a session-by-session basis. Nonlinear rupture trajectories were found on the individual level with high intra- and interindividual differences. In adolescent patients with BPD, ruptures tended to emerge in phases. Also Stevens et al. [29] found that 50% of patients (cluster C & NOS) showed local rupture resolution patterns. There was however no association to be found between the occurrence of rupture resolution patterns and outcome. On the other hand, Strauss et al. [26] found that 56% of their sample showed rupture resolution patterns. Higher pretreatment personality pathology (cluster C) resulted in less rupture resolution patterns ($r = .43$, $p < .05$). Occurrence of rupture resolution patterns in turn predicted improvement in personality pathology ($r = -.53$, $p < .01$) and depression ($r = .41$, $p < .05$).

When looking at resolutions, Muran et al. [27] found that patients (cluster C & NOS) who experience more resolutions of ruptures were less likely to drop out ($r = -.29$, $p < .05$). In contrast, for adolescents with BPD, Schenk et al. [•18, 30] showed that although dropouts experienced a higher frequency of alliance ruptures per session in comparison to the completers, therapists of the dropout patients applied an equal proportion of resolutions to ruptures as the therapists of the completers. This indicates, that the mere number of resolutions was not protective of dropping out in this particular sample. In the same vein, Gersh et al. [28] found that a higher number of resolutions in late sessions were associated with improvement in BPD symptoms ($r = -.67$, $p < .05$). Daly et al. [15] demonstrated that the complexity of resolution strategies was associated with the treatment outcome in a sample of three recovered and three unrecovered adolescent patients with BPD. Complexity was conceptualized by the number of stages of Bennet's [•14] resolution stage model employed by therapists. The resolution stage model was defined in a task analytical effort and engulfs nine hierarchical (sequential) stages, namely: acknowledgment, exploration, linking and explanation, negotiation, consensus, understanding and assimilating

warded off feelings, further explanation, change to patterns/aim, and closure. Daly et al. [15] showed that resolved rupture episodes are characterized by more and higher (negotiation to closure) model stages. Further, therapists of recovered patients used more and higher stage resolutions than therapists of unrecovered patients who more often remained at a lower stage of resolution. In contrast, Boritz et al. [17] found that withdrawal ruptures tended to persist over time in the unrecovered but not the recovered patients (BPD) despite the degree to which they were resolved in the prior session.

In summary, there is first meta-analytical evidence for the beneficial effect of rupture resolution processes during psychotherapy. When looking at PD literature, early alliance seems to influence dropout status and positively predict outcome. However, Strauss et al. [26] found that early alliance is associated with pretreatment pathology, indicating, that the beneficial effects of early alliance are found in patients that are able to build a stable therapeutic relationship in the early phase of therapy. In the same vein, the underlying principles of change in rupture resolution patterns still remain unclear given their possible link to pretreatment personality pathology. The studies concerned with resolutions and their complexity during the process point towards favorable effects of higher order resolutions and resolutions in the later phase of therapy. However, they also indicate that in some cases the amount and the complexity of resolutions remain futile. The contextual factors (therapist effects, patient interpersonal dysfunction) hindering the therapeutic effect of offered resolutions remains an important avenue for future research.

# 6 Rupture Resolution Training

Eubanks-Carter et al. [31] have developed an alliance focused training (AFT). Safran & Muran's publication on *Negotiating the Therapeutic Alliance: A Relational Treatment Guide* [8] serves as a training manual. Muran et al. [32] have tested the additive effect of AFT during a 30 session CBT protocol for cluster C PD on interpersonal behavior (SASB). Therapists were introduced to AFT at different time intervals (after session 8 or 16) controlling for patient, therapist and patient-therapist interactional effects. For patients, the introduction of AFT predicted a decrease in following behavior, an increase in expressiveness and a trend was found towards less avoidance and appeasement. For therapists, the onset of AFT training predicted a decrease in therapist blaming and directing and an increase in therapist affirmation and expressiveness. Findings

demonstrate that the expressing behavior of the patient was positively related to treatment outcome and patient appeasing and blaming were negatively related to treatment outcome. The authors insofar were able to show that AFT resulted in beneficial changes of interpersonal processes in a CBT setting.

Although the study of Muran et al. [32] demonstrates meaningful changes in interpersonal behavior, the impact of rupture resolution training on the outcome level remains sparse. The meta-analyses of Eubanks et al. [•9] did not support a significant effect of a rupture resolution training on the treatment outcome.

# 7 Need for new methods in future research

The finding that a better alliance is associated with a better treatment outcome has been confirmed in numerous studies [33-35]. However, in these study designs, the alliance is treated as a fixed effect factor that is measured most often at the beginning of the therapy. Therefore, the dynamic quality of the alliance is neglected. With advanced statistical methods and study designs the state- and trait-like components in the alliance-outcome association can be distinguished. This allows to analyze to what extent the alliance acts as a precondition to the therapeutic process (trait-like general relationship tendencies of the patient) or as a dynamic entity of the therapeutic relationship (i.e. state-like like rupture resolution processes) that operates as a vehicle of change [••21]. However, the video analysis of critical events is very labor- and cost-intensive. This results in studies with either low or no repeated measurement or in studies with small sample sizes. Future research should therefore invest in automated methods to describe alliance characteristics using physiological [36, 37], vocal [38, 39], facial [40] and movement [41] information streams. Interpersonal processes have been proposed to mirror the quality of interactions, especially synchrony has been proposed as an integrative framework for the therapeutic alliance [•42]. This theoretical framework allows to focus on objectively measurable interpersonal behavior and to complement the subjective experience and linguistic interaction studied so far in alliance research [43]. Alternatively, machine-learning could be used to approximate critical events in psychotherapy based on automated information streams and to mark them for further manual evaluation. For example, minimal response withdrawal ruptures could be approximated by automated silence detection [30] and speaker diarization [44]. The extent of nonverbal movement

synchrony has been associated with patient rated relationship quality [41], dropout status and improvement [45]. Reich et al. [39] found that vocal synchrony in turn was negatively associated with the patient rated working alliance. The here proposed methods assess alliance ruptures only indirectly but on a fine-grained temporal level in large samples sizes over the whole psychotherapeutic process. This would allow to study bridging concepts like synchrony or emotion regulation in relation to working alliance processes for PD specifically.

# 8 Conclusion

Alliance ruptures and their resolution in PD are studied through direct (observer-based methods: 3RS, CIS) and indirect methods (e.g. questionnaires: PSQ, WAI). Horvath [46] estimates over 70 different instruments that operationalize the alliance based on different theoretical constructs. The biggest problems of the field are objectivity and comparability. Studies are sparse and highly heterogeneous (PD cluster, research aims, session sampling, methods), impeding the comparability and interpretation of findings. Concordance of direct and indirect measures of alliance ruptures must be considered low [47, 19]. In order to avoid a forest of unconnected and therefore hard to summarize findings, methodological objectivity should be considered a key interest for the field.

Throughout the literature there is frequent use of cluster C patient samples, cluster A and B studies are highly encouraged. It is hypothesized that PD diagnoses influence the quality of ruptures encountered through their characteristic interpersonal constraints [•20, ••3]. It is therefore important to acquire and compare findings for all different types of PD. Moving towards the dimensional ICD-11 diagnosis, future studies should include dimensional personality functioning as covariate variables to embed findings. Results by Strauss et al. [26] or Muran et al. [27] already point towards the direction that personality functioning is a more feasible moderator and outcome variable than the number of SCID items. Further, few studies test for trait-like pretreatment factors moderating the alliance outcome association. Dimensional pretreatment personality functioning (DSM-5; [•48, •49]) is considered an important variable for future studies, in order to report what has worked for whom.

Table 1: Direct and indirect instruments of alliance ruptures

| | Instruments | Perspective | Details |
|---|---|---|---|
| Direct Instruments | Rupture Resolution Rating System (3RS; [22]) | Observer Rating | Contains seven withdrawal and seven confrontation rupture markers as well as ten resolution markers. Markers are rated on a significance rating scale and an overall resolution rating scale. Unit of coding are speaking turns, 1- or 5-minute windows or defining start and stop markers for each episode. |
| | Collaborative Interaction Scale-Revised (CIS-R ; [23]) | Observer Rating | 29-item rating scale with two subscales CIS-P and CIS-T. CIS-P defines direct and indirect rupture markers and direct and indirect collaborative processes of the patient. CIS-T defines direct and indirect collaborative interventions, rupture and therapist interventions. Coding is done within narrative units. Ruptures and resolutions can be coded for both the patient and the therapist. |
| | Harper's [50, 51] manual of rupture markers | Observer Rating | Defines ten markers of confrontation ruptures and eight markers of withdrawal ruptures. The analysis is performed on the level of speaking turns. |
| | The Structural Analysis of Social Behavior (SASB; [52]) | Observer Rating | Measures interpersonal behavior of the patient and therapist with the octants "focus on other" and "focus on self" with two orthogonal dimensions interdependence (from autonomy to involvement) and affiliation (from hostility to friendliness). Alliance ruptures were rated as patient behaviors of appeasing, avoiding and blaming. Resolutions were rated as patient and therapist expressing and therapist affirming and directing [●9]. |
| Direct & indirect Instruments | Post-Session-Questionnaire [53] | Patient, Therapist | Combines the session evaluation questionnaire (SEQ; [54]) and *direct* questions of whether or not there was a problematic event in the relationship during the session and when these events occurred. Patients rate the tension of the problematic event on a 5-point likert scale. The SEQ assesses the session impact (subscales depth and smoothness) and post-session feelings of positivity and arousal. |
| Indirect Instruments | Working Alliance Inventory Short Form (WAI-SF; [55]) | Patient, Therapist, Observer | 12-item self-report with three subscales goals, tasks and bond of the therapeutic alliance. It is mostly administered at the end of sessions. |
| | California Psychotherapy Alliance Scales (CALPAS; [56]) | Patient, Therapist, Observer | The subscales patient working capacity, patient commitment, working strategy consensus and therapist understanding and involvement are assessed. Observer rating is done on the session level. |
| | Vanderbilt Therapeutic Alliance Scale-Revised Short Form (VTAS-R SF; [57]) | Observer | Five-item rating scale with two subscales patient contribution and patient-therapist interaction. The items reflect Bordin's theoretical constructs bond, goals and tasks. Items are scored on a 6-point Likert-Scale. The coding is done on session level. |
| | Alliance Negotiation Scale (ANS; [58]) | Patient | 12-item self-report scale that assesses the degree of constructive negotiation of disagreements about tasks and goals from the perspective of the patient. |
| | The Agnew Relationship Measure (ARM; [59]) | Patient, Therapist | 26-item questionnaire measuring the quality of the therapeutic relationship from the perspective of the therapist and the patient. |

**References**

● of importance, ●● of major importance

1. Ackerman SJ, Benjamin LS, Beutler LE, Gelso CJ, Goldfried MR, Hill C et al. Empirically supported therapy relationships: Conclusions and recommendations fo the Division 29 Task Force. Psychotherapy: Theory, Research, Practice, Training. 2001;38(4):495-7. doi:10.1037/0033-3204.38.4.495.

2. Norcross JC. Psychotherapy relationships that work: Therapist contributions and responsiveness to patients. Oxford University Press.; 2002.

3. ●●Tufekcioglu S, Muran JC. A relational approach to personality disorder & alliance rupture. Integrated treatment for personality disorders. New York: Guilford; 2014. **Presenting a theoretical link between PD and alliance**

4. Safran JD. Breaches in the Therapeutic Alliance - an Arena for Negotiating Authentic Relatedness. Psychotherapy. 1993;30(1):11-24. doi:Doi 10.1037/0033-3204.30.1.11.

5. Safran JD, Muran JC. The resolution of ruptures in the therapeutic alliance. J Consult Clin Psychol. 1996;64(3):447-58. doi:10.1037//0022-006x.64.3.447.

6. National Institute for Health and Care Excellence. Borderline personality disorder: recognition and management (NICE guideline 78). London: Department of Health; 2009. https://www.nice.org.uk/guidance/cg78

7. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.; 2013.

8. Safran JD, Muran JC. Negotiating the therapeutic alliance: A relational treatment guide. New York: Guilford Press; 2000.

9. •Eubanks CF, Muran JC, Safran JD. Alliance rupture repair: A meta-analysis. Psychotherapy (Chic). 2018;55(4):508-19. doi:10.1037/pst0000185. **Meta-analyses on the effects of rupture-resolution and resolution trainings on patient outcomes.**

10. Muran JC. Confessions of a New York rupture researcher: An insider's guide and critique. Psychother Res. 2019;29(1):1-14. doi:10.1080/10503307.2017.1413261.

11. Eubanks-Carter C, Gorman BS, Muran JC. Quantitative naturalistic methods for detecting change points in psychotherapy research: an illustration with alliance ruptures. Psychother Res. 2012;22(6):621-37. doi:10.1080/10503307.2012.693772.
12. Greenberg LS. Change process research. J Consult Clin Psychol. 1986;54(1):4-9. doi:10.1037//0022-006x.54.1.4.
13. Aspland H, Llewelyn S, Hardy GE, Barkham M, Stiles W. Alliance ruptures and rupture resolution in cognitive-behavior therapy: a preliminary task analysis. Psychother Res. 2008;18(6):699-710. doi:10.1080/10503300802291463.
14. •Bennett D, Parry G, Ryle A. Resolving threats to the therapeutic alliance in cognitive analytic therapy of borderline personality disorder: a task analysis. Psychol Psychother. 2006;79(Pt 3):395-418. doi:10.1348/147608305X58355. **Task analytical model of resolution in Cognitive Analytic Therapy providing in-depth description of resolution efforts.**

15. Daly AM, Llewelyn S, McDougall E, Chanen AM. Rupture resolution in cognitive analytic therapy for adolescents with borderline personality disorder. Psychol Psychother. 2010;83(Pt 3):273-88. doi:10.1348/147608309X481036.
16. Coutinho J, Ribeiro E, Fernandes C, Sousa I, Safran JD. The development of the therapeutic alliance and the emergence of alliance ruptures. [El desarrollo de la alianza terapéutica y la aparición de rupturas en la alianza]. Anales de Psicología. 2014;30(3). doi:10.6018/analesps.30.3.168911.
17. Boritz T, Barnhart R, Eubanks CF, McMain S. Alliance Rupture and Resolution in Dialectical Behavior Therapy for Borderline Personality Disorder. J Pers Disord. 2018;32(Suppl):115-28. doi:10.1521/pedi.2018.32.supp.115.
18. •Schenk N, Zimmermann R, Fürer L, Krause M, Weise S, Kaess M et al. Trajectories of alliance ruptures in the psychotherapy of adolescents with borderline personality pathology: timing, typology and significance. Research in Psychotherapy: Psychopathology, Process and Outcome. 2019;22(2). doi:10.4081/ripppo.2019.348. **Presents graphics of alliance trajectories in adolescents with a borderline personality pathology. Provides a session-to-session measurement of complete psychotherapies with the 3RS.**

19. Sommerfeld E, Orbach I, Zim S, Mikulincer M. An in-session exploration of ruptures in working alliance and their associations with clients' core conflictual relationship themes, alliance-

related discourse, and clients' postsession evaluations. Psychotherapy Research. 2008;18(4):377-88. doi:10.1080/10503300701675873.

20. •Bender DS. The therapeutic alliance in the treatment of personality disorders. J Psychiatr Pract. 2005;11(2):73-87. doi:10.1097/00131746-200503000-00002. **Clinical considerations and theory about the link between the interpersonal nature of Personality Disorders and rupture resolution**

21. ••Zilcha-Mano S. Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. Am Psychol. 2017;72(4):311-25. doi:10.1037/a0040435. **Introduces the differentiation between trait-like and state-like components in alliance and bringing to important new avenues for alliance research.**

22. Eubanks CF, Lubitz J, Muran JC, Safran JD. Rupture Resolution Rating System (3RS): Development and validation. Psychother Res. 2019;29(3):306-19. doi:10.1080/10503307.2018.1552034.

23. Colli A, Gentile D, Condino V, Lingiardi V. Assessing alliance ruptures and resolutions: Reliability and validity of the Collaborative Interactions Scale-revised version. Psychother Res. 2019;29(3):279-92. doi:10.1080/10503307.2017.1414331.

24. Tufekcioglu S, Muran JC, Safran JD, Winston A. Personality disorder and early therapeutic alliance in two time-limited therapies. Psychother Res. 2013;23(6):646-57. doi:10.1080/10503307.2013.843803.

25. Safran JD, Muran JC, Eubanks-Carter C. Repairing alliance ruptures. Psychotherapy (Chic). 2011;48(1):80-7. doi:10.1037/a0022140.

26. Strauss JL, Hayes AM, Johnson SL, Newman CF, Brown GK, Barber JP et al. Early alliance, alliance ruptures, and symptom change in a nonrandomized trial of cognitive therapy for avoidant and obsessive-compulsive personality disorders. J Consult Clin Psychol. 2006;74(2):337-45. doi:10.1037/0022-006X.74.2.337.

27. Muran JC, Safran JD, Gorman BS, Samstag LW, Eubanks-Carter C, Winston A. The relationship of early alliance ruptures and their resolution to process and outcome in three time-limited psychotherapies for personality disorders. Psychotherapy (Chic). 2009;46(2):233-48. doi:10.1037/a0016085.

28. Gersh E, Hulbert CA, McKechnie B, Ramadan R, Worotniuk T, Chanen AM. Alliance rupture and repair processes and therapeutic change in youth with borderline personality disorder. Psychol Psychother. 2017;90(1):84-104. doi:10.1111/papt.12097.
29. Stevens CL, Muran JC, Safran JD, Gorman BS, Winston A. Levels and patterns of the therapeutic alliance in brief psychotherapy. Am J Psychother. 2007;61(2):109-29. doi:10.1176/appi.psychotherapy.2007.61.2.109.
30. Zimmermann R, Furer L, Schenk N, Koenig J, Roth V, Schluter-Muller S et al. Silence in the psychotherapy of adolescents with borderline personality pathology. Personal Disord. 2020. doi:10.1037/per0000402.
31. Eubanks-Carter C, Muran JC, Safran JD. Alliance-focused training. Psychotherapy (Chic). 2015;52(2):169-73. doi:10.1037/a0037596.
32. Muran JC, Safran JD, Eubanks CF, Gorman BS. The effect of alliance-focused training on a cognitive-behavioral therapy for personality disorders. J Consult Clin Psychol. 2018;86(4):384-97. doi:10.1037/ccp0000284.
33. Fluckiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic). 2018;55(4):316-40. doi:10.1037/pst0000172.
34. Horvath AO, Del Re AC, Fluckiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy (Chic). 2011;48(1):9-16. doi:10.1037/a0022186.
35. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol. 2000;68(3):438-50.
36. Kleinbub JR. State of the Art of Interpersonal Physiology in Psychotherapy: A Systematic Review. Front Psychol. 2017;8:2053. doi:10.3389/fpsyg.2017.02053.
37. Marci CD, Orr SP. The effect of emotional distance on psychophysiologic concordance and perceived empathy between patient and interviewer. Appl Psychophysiol Biofeedback. 2006;31(2):115-28. doi:10.1007/s10484-006-9008-4.
38. Imel ZE, Barco JS, Brown HJ, Baucom BR, Baer JS, Kircher JC et al. The association of therapist empathy and synchrony in vocally encoded arousal. J Couns Psychol. 2014;61(1):146-53. doi:10.1037/a0034943.
39. Reich CM, Berman JS, Dale R, Levitt HM. Vocal Synchrony in Psychotherapy. Journal of Social and Clinical Psychology. 2014;33(5):481-94. doi:10.1521/jscp.2014.33.5.481.

40. Steppan M, Zimmermann R, Fürer L, Schenk N, Schmeck K. Machine Learning Facial Emotion Recognition in Psychotherapy Research. A useful approach?. PsyArXiv. 2020.
41. Ramseyer F, Tschacher W. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. J Consult Clin Psychol. 2011;79(3):284-95. doi:10.1037/a0023419.
42. •Koole SL, Tschacher W. Synchrony in Psychotherapy: A Review and an Integrative Framework for the Therapeutic Alliance. Front Psychol. 2016;7:862. doi:10.3389/fpsyg.2016.00862. **Reviews the course of alliance research and introducing synchrony as a framework for alliance.**

43. Tschacher W, Haken H, Kyselo M. Alliance: a common factor of psychotherapy modeled by structural theory. Front Psychol. 2015;6:421. doi:10.3389/fpsyg.2015.00421.
44. Fürer L, Schenk N, Roth V, Steppan M, Schmeck K, Zimmermann R. Supervised speaker diarization using random forests: A tool for psychotherapy process research. Frontiers in Psychology. accepted.
45. Paulick J, Deisenhofer AK, Ramseyer F, Tschacher W, Boyle K, Rubel J et al. Nonverbal Synchrony: A New Approach to Better Understand Psychotherapeutic Processes and Drop-Out. Journal of Psychotherapy Integration. 2018;28(3):367-84. doi:10.1037/int0000099.
46. Horvath AO. Research on the alliance: Knowledge in search of a theory. Psychother Res. 2018;28(4):499-516. doi:10.1080/10503307.2017.1373204.
47. Coutinho J, Ribeiro E, Sousa I, Safran JD. Comparing two methods of identifying alliance rupture events. Psychotherapy (Chic). 2014;51(3):434-42. doi:10.1037/a0032171.
48. •Goth K, Birkholzer M, Schmeck K. Assessment of Personality Functioning in Adolescents With the LoPF-Q 12-18 Self-Report Questionnaire. J Pers Assess. 2018;100(6):680-90. doi:10.1080/00223891.2018.1489258. **Describes the LoPF Questionnaire for the assessment of personality functioning in adolescents.**

49. •Hutsebaut J, Kamphuis JH, Feenstra DJ, Weekers LC, De Saeger H. Assessing DSM-5-oriented level of personality functioning: Development and psychometric evaluation of the Semi-Structured Interview for Personality Functioning DSM-5 (STiP-5.1). Personal Disord. 2017;8(1):94-101. doi:10.1037/per0000197. **Describes the semi-structured interview for the assessment of personality functioning in adults.**

50. Harper H. Coding Guide I: Identification of confrontation challenges in exploratory therapy. In: Sheffield Uo, editor.1989.

51. Harper H. Coding Guide II: Identification of withdrawal challenges in exploratory therapy. In: Sheffield Uo, editor.1989.

52. Benjamin LS. SASB - a Bridge between Personality Theory and Clinical-Psychology. Psychological Inquiry. 1994;5(4):273-316. doi:DOI 10.1207/s15327965pli0504_1.

53. Samstag LW, Batchelder ST, Muran JC, Safran JD, Winston A. Early identification of treatment failures in short-term psychotherapy. An assessment of therapeutic alliance and interpersonal behavior. J Psychother Pract Res. 1998;7(2):126-43.

54. Stiles WB. Measurement of the Impact of Psychotherapy Sessions. Journal of Consulting and Clinical Psychology. 1980;48(2):176-85. doi:Doi 10.1037/0022-006x.48.2.176.

55. Horvath AO, Greenberg LS. Development and Validation of the Working Alliance Inventory. Journal of Counseling Psychology. 1989;36(2):223-33. doi:Doi 10.1037/0022-0167.36.2.223.

56. Marmar CR, Gaston L. California Psychotherapy Alliance Scales (CALPAS) manual. In: California Uo, editor. Unpublished manuscript1988.

57. Shelef K, Diamond GM. Short form of the revised Vanderbilt therapeutic alliance scale: development, reliability, and validity. Psychother Res. 2008;18(4):433-43. doi:10.1080/10503300701810801.

58. Doran JM, Safran JD, Waizmann V, Bolger K, Muran JC. The alliance negotiation scale: psychometric construction and preliminary reliability and validity analysis. Psychother Res. 2012;22(6):710-9. doi:10.1080/10503307.2012.709326.

59. Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). Br J Clin Psychol. 1998;37(2):155-72. doi:10.1111/j.2044-8260.1998.tb01291.x.

## Results

I shortly summarize the results of the research papers.

**S1:** We showed that a supervised method for speaker diarization is feasible for psychotherapy process research, offering a good balance between manual workload and error management. Errors were substantially lower than with unsupervised methods and comparable to other supervised procedures. Further, errors were higher in same sex dyads than different sex dyads. By using the random forest algorithm, it was possible to attain a measure for the quality of diarization in a dyad, based only on a learning set (out-of-bag error). These results show that supervised speaker diarization based on our features set is a robust method practicable to be used in future research towards vocal features and conversational parameters.

**S2:** We found silence episodes (> 3 seconds) in all sessions. On average silence comprises around 10% of the total session time. We did not find a link between pre-treatment psychopathology (personality functioning and depression) and silence. However, sessions with silence were perceived more negatively by patients. They were rated less good and less smooth (unpleasant, rough, difficult, uncomfortable). We concluded that the use of silence as a technique is not well tolerated in the investigated sample.

**S3:** We found that the method used for machine learning facial emotion recognition showed substantial, but numerically low agreement with the human ratings of affect. The numerically low agreement has been discussed in light of mismatch between the rating method and the automatic emotion recognition: human rated overall affect in one-minute windows, selecting only one specific emotion per window, based on all possible verbal and nonverbal streams of information. In a minute window, more than one (facial) emotion is likely to occur, therefor results are constrained by heavy averaging. However, results showed high objectivity, being equal in strength across the three different raters. Concerning the outcome analysis, a strong effect indicated the beneficial effect of happiness (Bänninger-Huber, 1992). Although the study underlies statistical constraints, it can be concluded, that the method offers a reliable and valid tool for psychotherapy process research.

**S4:** The review conducted summarized research on the rupture repair process in personality disorders. Reviewing efforts were substantially hindered by sparse and highly heterogeneous data (PD cluster, research aims, session sampling, methods). Especially studies on cluster A and B personality disorders are highly encouraged. The literature indicates that ruptures are very common in patients with personality disorders (Eubanks et al., 2019; Schenk et al., 2019, 2020). Further, we found evidence indicating that ruptures are more frequent in personality disorders than other clinical samples and that there are differences in therapist behaviour with mentioned sample comparison. We conclude that lower interpersonal functioning in these patients endangers the building and maintaining of the therapeutic relationship. Although evidence indicates that rupture resolution positively influences dropout status and outcome, some studies also find a correlation between early alliance and pre-treatment pathology. This indicates that early alliance may not be due to the therapeutic process (Zilcha-Mano, 2016, 2017, 2019; Zilcha-Mano et al., 2014; Zilcha-Mano & Errázuriz, 2017), but instead may be dependent on pre-therapeutic experiences. Accordingly, the underlying principle of change in rupture resolution processes has not been studied rigorously enough to interpret outcome in light of the occurrence of rupture repairs. Further, future studies are advised to include dimensional personality functioning to discuss what works for whom (Norcross & Wampold, 2011).

## Discussion

Short term future directions, close to the methods and findings in the papers, have already been addressed in the publications. Here I summarize long term goals that further the study of emotion regulation and alliance negotiation in the framework of the events paradigm and the BPD relevant change mechanism towards identity integration. Overall, automated analyses of nonverbals hold the promise of providing therapists with key therapeutic moments of therapy that can later be discussed in supervision or be revisited before the next session, guiding the therapists´ thought process.

Theory on identity integration posits that the psychotherapeutic work involves affect activation and may be accompanied by interpersonal defences, i.e., ruptures (Greenberg & Safran, 1989; Yeomans et al., 2013). Finding nonverbal patterns indicating both affect activation episodes and strains in the therapeutic relationship is pivotal.

Automatic recognition of affect activation has been achieved using machine learning facial emotion recognition (Steppan et al., submitted). Although stable across individuals, the correlation with overall human rated arousal is small ($r=-.104$, $p<6.46e-24$). This could stem from the fact, that both systems were not adapted on the same time scale. In a next step, it would be feasible to define moments of affect activation longer than one-minute lengths from the human ratings and test whether the output of facial emotion recognition software is able to predict these longer episodes. This would deliver a first delineation of moments of interest. Also, multimodal affect recognition is standard nowadays (Poria et al., 2017), future efforts should include arousal measures from the voice. The voice holds one crucial benefit: From a clinical standpoint, therapy is a private endeavour where patient comfort is crucial. Rooms with cameras may stand in contrast to the safe space therapy should provide. The voice can be captured in a non-invasive manner, but still revealing a lot in terms of features. This stream may be more promising for future everyday adaptations in the clinical field.

Strains in the therapeutic relationship in terms of ruptures have already been assessed in the here used data (Schenk et al., 2019, 2020). Future studies could concentrate on the description of rupture and resolution episodes in terms of affect (measured in voice and facial expression) or in terms of conversational parameters as silence, turn taking behaviour or rhythm, implying disfluencies in the ongoing negotiation of the therapeutic dyad.

For the specific study of identity integration, it would be favourable to include qualitative assessment of in-session enactments of object relations using the object relations rating scale (Diguer et al., 2012). Here, automatically assessed nonverbals could be anchored to further understand their use in the process of identity integration.

This dissertation has been working under the premise, that theory driven meaningful patterns of nonverbal interaction can be deducted and studied using psychotherapeutic theory in relation to the therapy of borderline personality disorders. Another approach is that the process of meaning making is outsourced to the users (therapists). In that sense, the task of future therapy applications would be to visualize concepts of transtheoretical interest such as affect activation and conversational disfluencies (silences). The interpretation thereof then lies in the hands of therapists and their related school of thought.

Concerning the problem of interindividual (patient and therapist) and dyadic differences in the use of nonverbals (Crangle et al., 2019; Poria et al., 2017), future clinical applications could be advanced using self-organisation and complexity analyses in order to find dyad specific key moments in the interactive process, rather than specific combinations of feature expressions linked to affect (Schiepek & Strunk, 2010).

Both affect activation and strains in the therapeutic relationship have been proposed to both be (in part) nonverbal acts in an ongoing self-state sharing. This ignores the semantic aspect of the verbal interaction. In order to complement the study of nonverbal interactions, automatic speech recognition may provide therapists with transcripts and thematic analyses and word use analyses in the coming decades. English speaking countries already profit from open-source applications (Halfon et al., 2020), swiss German speech recognition is in the making and will soon evolve from strictly commercial use (Hurry et al., 2020).

How automated variables will be clinically used in the future is still unknown. Methods in affective computing are fast changing and en vogue right now. It is likely that the field will keep evolving and that methods will improve in the coming years. As such it is important to start integrating findings to more robust evidence. For example, until now, only a hand full of studies concern with automated audio analyses in the field of psychotherapy (Imel et al., 2014; Reich et al., 2014; Rochman & Amir, 2013; Soma et al., 2020; Xiao et al., 2015). I therefor think it should be considered to work on providing convenient tools (speaker diarization) to the research community to perform automated analyses on a bigger scale (Fürer, 2020). It is of high importance that the methods do not stay in the hands of a selected few, but rather, that the gathering and interpretation of findings can be distributed to various research groups working with diverse patients in differing settings.

## References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2020). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, 1–25. https://doi.org/10.1080/10503307.2020.1808729

Arango, I., Miranda, E., Sánchez Ferrer, J. C., Fresán, A., Reyes Ortega, M. A., Vargas, A. N.,

Barragán, S., Villanueva-Valle, J., Munoz-Delgado, J., & Robles, R. (2019). Changes in facial emotion expression during a psychotherapeutic intervention for patients with borderline personality disorder. *Journal of Psychiatric Research*, *114*, 126–132. https://doi.org/10.1016/j.jpsychires.2019.04.026

Armstrong, C. R., Tanaka, S., Reoch, L., Bronstein, L., Honce, J., Rozenberg, M., & Powell, M. A. (2015). Emotional arousal in two drama therapy core processes: Dramatic embodiment and dramatic projection. *Drama Therapy Review*, *1*(2), 147–160.

Auszra, L., Greenberg, L. S., & Herrmann, I. (2013). Client emotional productivity—Optimal client in-session emotional processing in experiential therapy. *Psychotherapy Research*, *23*(6), 732–746. https://doi.org/10.1080/10503307.2013.816882

Baldwin, M. W. (1992). Relational Schemas and the Processing of Social Information. *Psychological Bulletin*, *112*(3), 461–484.

Bänninger-Huber, E. (1992). Prototypical Affective Microsequences in Psychotherapeutic Interaction. *Psychotherapy Research*, *2*(4), 291–306. https://doi.org/10.1080/10503309212331333044

Beebe, B., Lachmann, F., & Jaffe, J. (1997). Mother—Infant interaction structures and presymbolic self- and object representations. *Psychoanalytic Dialogues*, *7*(2), 133–182. https://doi.org/10.1080/10481889709539172

Beebe, B., & Lachmann, F. M. (1998). Co-constructing inner and relational processes: Self and mutual regulation in infant research and adult treatment. *Psychoanalytic Psychology*, *Vol.15*(4), 480–516.

Benzi, I., & Madeddu, F. (2017). Development of Personality Disorders: Identity as a Key Process. *International Journal of Psychology & Behavior Analysis*, *3*(1). https://doi.org/10.15344/2455-3867/2017/124

Bernieri, F., J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. *In R. Feldman & B. Rimé (Eds.), Fundamentals of Nonverbal Behavior*, *New York: Cambridge University Press.*, (pp. 401–432).

Birkhölzer, M., Goth, K., Schrobildgen, C., Schmeck, K., & Schlüter-Müller, S. (2015). Grundlagen und praktische Anwendung des Assessments of Identity Development in Adolescence (AIDA)/ Background and Practical Use of the Assessment of Identity Development in Adolescence

(AIDA). *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *64*(8), 584–600. https://doi.org/10.13109/prkk.2015.64.8.584

Birkhölzer, M., Schlüter-Müller, S., & Schmeck, K. (2020). Persönlichkeitsstörungen im Kindes- und Jugendalter. *Swiss Archives of Neurology, Psychiatry and Psychotherapy*, 6. https://doi.org/10.4414/sanp.2020.03105

Böhmer, M. (2010). "Communication by impact" and other forms of non-verbal communication: A review of transference, counter- transference and projective identification. *African Journal of Psychiatry*, *13*, 179–183.

Bond, M., Paris, J., & Zweig-Frank, H. (1994). Defense styles and borderline personality disorder. *Journal of Personality Disorders*, *8*(1), 28–31.

Bone, D., Lee, C.-C., & Narayanan, S. (2014). Robust Unsupervised Arousal Rating:A Rule-Based Framework withKnowledge-Inspired Vocal Features. *IEEE Transactions on Affective Computing*, *5*(2), 201–213. https://doi.org/10.1109/TAFFC.2014.2326393

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.

Brokerhof, I. M., Bal, P. M., Jansen, P. G. W., & Solinger, O. N. (2018). Ficitional Narratives and Identity Change: Three pathways through which stories influence the dialogical self. *Dialogical Self: Inspirations, Considerations and Research*, 29–57.

Buchholz, M. B. (2019). Psychodynamic Psychotherapy Process Research: A quick ride through what you should know about process (research). *Journal of Analytical Psychology*, *64*(5), 798–822. https://doi.org/10.1111/1468-5922.12547

Campos, J. J., Walle, E. A., Dahl, A., & Main, A. (2011). Reconceptualizing Emotion Regulation. *Emotion Review*, *3*(1), 26–35. https://doi.org/10.1177/1754073910380975

Carryer, J. R., & Greenberg, L. S. (2010). Optimal levels of emotional arousal in experiential therapy of depression. *Journal of Consulting and Clinical Psychology*, *78*(2), 190–199. https://doi.org/10.1037/a0018401

Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: A task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, 8.

Centonze, A., Inchausti, F., MacBeth, A., & Dimaggio, G. (2020). Changing Embodied Dialogical Patterns in Metacognitive Interpersonal Therapy. *Journal of Constructivist Psychology*, 1–15. https://doi.org/10.1080/10720537.2020.1717117

Charny, J. E. (1966). Psychosomatic Manifestations of Rapport in Psychotherapy: *Psychosomatic Medicine*, *28*(4), 305–315. https://doi.org/10.1097/00006842-196607000-00002

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Clarkin, J. F., Lenzenweger, M. F., Yeomans, F., Levy, K. N., & Kernberg, O. F. (2007). An Object Relations Model of Borderline Pathology. *Journal of Personality Disorders*, *21*(5), 474–499. https://doi.org/10.1521/pedi.2007.21.5.474

Crangle, C. E., Wang, R., Perreau-Guimaraes, M., Nguyen, M. U., Nguyen, D. T., & Suppes, P. (2019). Machine learning for the recognition of emotion in the speech of couples in psychotherapy using the Stanford Suppes Brain Lab Psychotherapy Dataset. *ArXiv Preprint*, *arXiv:1901.04110*, 24.

Cuttler, E., Hill, C. E., King, S., & Kivlighan, D. M. (2019). Productive silence is golden: Predicting changes in client collaboration from process during silence and client attachment style in psychodynamic psychotherapy. *Psychotherapy*, *56*(4), 568–576. https://doi.org/10.1037/pst0000260

D'Agostino, A., Covanti, S., Rossi Monti, M., & Starcevic, V. (2017). Reconsidering Emotion Dysregulation. *Psychiatric Quarterly*, *88*(4), 807–825. https://doi.org/10.1007/s11126-017-9499-6

Dales, S., & Jerry, P. (2008). Attachment, Affect Regulation and Mutual Synchrony in Adult Psychotherapy. *American Journal of Psychotherapy*, *62*(3), 283–312. https://doi.org/10.1176/appi.psychotherapy.2008.62.3.283

De Meulemeester, C., Lowyck, B., Vermote, R., Verhaest, Y., & Luyten, P. (2017). Mentalizing and interpersonal problems in borderline personality disorder: The mediating role of identity diffusion. *Psychiatry Research*, *258*, 141–144. https://doi.org/10.1016/j.psychres.2017.09.061

Diguer, L., Gamache, D., & Laverdière, O. (2012). Development and initial validity of the Object Relations Rating Scale. *Psychotherapy Research*, *22*(4), 402–416. https://doi.org/10.1080/10503307.2012.662606

Dosamantes-Alperson, E. (1979). The intrapsychic and the interpersonal in movement psychotherapy. *American Journal of Dance Therapy*, *3*(1), 20–31. https://doi.org/10.1007/BF02579546

Dreyer, F. (2018). „Und ich bin der Räuber“: Erwartungsstrukturen in der alternativen Realität des Spiels. *Paragrana*, *27*(1), 237–252. https://doi.org/10.1515/para-2018-0017

Eigen, M. (1977). On breathing and identity. *Journal of Humanistic Psychology*, *17*(3), 35–39. https://doi.org/10.1177/002216787701700304

Elliott, R. (2010). Psychotherapy change process research: Realizing the promise. *Psychotherapy Research*, *20*(2), 123–135. https://doi.org/10.1080/10503300903470743

Eubanks, C. F., Lubitz, J., Muran, J. C., & Safran, J. D. (2019). Rupture Resolution Rating System (3RS): Development and validation. *Psychotherapy Research*, *29*(3), 306–319. https://doi.org/10.1080/10503307.2018.1552034

Feenstra, D. J., Hutsebaut, J., Verheul, R., & van Limbeek, J. (2014). Identity: Empirical Contribution: Changes in the Identity Integration of Adolescents in Treatment for Personality Disorders. *Journal of Personality Disorders*, *28*(1), 101–112. https://doi.org/10.1521/pedi.2014.28.1.101

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. American Psychiatric Press.

Foelsch, P. A., Odom, A., Arena, H., Krischer, M. K., Schmeck, K., & Schlüter-Müller, S. (2010). The Differentiation between Identity Crisis and Identity Diffusion and its Impact on Treatment. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *59*, 418–434.

Foelsch, P. A., Schlüter-Müller, S., Odom, A. E., Arena, H. T., Borzutzky H., A., & Schmeck, K. (2014). *Adolescent Identity Treatment*. Springer International Publishing. https://doi.org/10.1007/978-3-319-06868-8

Foubert, K. (2017). Impaired Maintenance of Interpersonal Synchronization in Musical Improvisations of Patients with Borderline Personality Disorder. *Frontiers in Psychology*, *8*, 17.

Foubert, K. (2018). *Transformation in Improvisation*. KU Leuven.

Foubert, K., Sebreghts, B., Sutton, J., & De Backer, J. (2020). Musical encounters on the borderline. Patterns of mutuality in musical improvisations with Borderline Personality Disorder. *The*

*Arts in Psychotherapy*, *67*, 8. https://doi.org/10.1016/j.aip.2019.101599

Fürer, L. (2020). *Supervised Dyadic Speaker Diarization* (Version v1) [Computer software]. GitHub. https://github.com/lukasfuerer/Supervised-Dyadic-Speaker-Diarization

Gagnon, J., Vintiloiu, A., & McDuff, P. (2016). Do splitting and identity diffusion have respective contributions to borderline impulsive behaviors? Input from Kernberg's model of personality. *Psychoanalytic Psychology*, *33*(3), 420–436. https://doi.org/10.1037/pap0000052

Geiger, P. J., Peters, J. R., & Baer, R. A. (2014). Using a measure of cognitive distortion to examine the relationship between thought suppression and borderline personality features: A multi-method investigation. *Personality and Individual Differences*, *59*, 54–59. https://doi.org/10.1016/j.paid.2013.11.005

Goldberg, S. B., Flemotomos, N., Martinez, V. R., Tanana, M. J., Kuo, P. B., Pace, B. T., Villatte, J. L., Georgiou, P. G., Epps, J. V., Imel, Z. E., Narayanan, S. S., & Atkins, D. C. (2020). Machine Learning and Natural Language Processing in Psychotherapy Research: Alliance as Example Use Case. *Journal of Counseling Psychology*, *Vol. 67*(No. 4), 438–448.

Golynkina, K., & Ryle, A. (1999). The identification and characteristics of the partially dissociated states of patients with borderline personality disorder. *British Journal of Medical Psychology*, *72*(4), 429–445. https://doi.org/10.1348/000711299160103

Goth, K., Birkhölzer, M., & Schmeck, K. (2018). Assessment of Personality Functioning in Adolescents With the LoPF–Q 12–18 Self-Report Questionnaire. *Journal of Personality Assessment*, *100*(6), 680–690. https://doi.org/10.1080/00223891.2018.1489258

Goth, K., Foelsch, P., Schlüter-Müller, S., Birkhölzer, M., Jung, E., Pick, O., & Schmeck, K. (2012). Assessment of identity development and identity diffusion in adolescence—Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child and Adolescent Psychiatry and Mental Health*, *6*(1), 27. https://doi.org/10.1186/1753-2000-6-27

Grace, M., Kivlighan Jr, D. M., & Kunce, J. (1995). The effect of nonverbal skills training on counselor trainee nonverbal sensitivity and responsiveness and on session impact and working alliance ratings. *Journal of Counseling & Development*, *73*(5), 547–552.

Greenberg, L. S. (1991). Research on the process of change. *Psychotherapy Research*, *1*(1), 3–

16. https://doi.org/10.1080/10503309112331334011

Greenberg, L. S., & Paivio, S. C. (2003). *Working with emotions in psychotherapy: Vol. Vol. 13*. Guilford Press.

Greenberg, L. S., & Pascual-Leone, A. (2006). Emotion in psychotherapy: A practice-friendly research review. *Journal of Clinical Psychology*, *62*(5), 611–630. https://doi.org/10.1002/jclp.20252

Greenberg, L. S., & Safran, J. D. (1989). Emotion in Psychotherapy. *American Psychologist*, *44*(1), 19–29.

Halfon, S., Doyran, M., Türkmen, B., Oktay, E. A., & Salah, A. A. (2020). Multimodal affect analysis of psychodynamic play therapy. *Psychotherapy Research*, 1–16. https://doi.org/10.1080/10503307.2020.1839141

Håvås, E., Svartberg, M., & Ulvenes, P. (2014). Attuning to the unspoken: The Relationship Between Therapist Nonverbal Attunement and Attachment Security in Adult Psychotherapy. *Psychoanalytic Psychology*, 21. http://dx.doi.org/10.1037/a0038517

Hermans, H. J. M. (2006). The Self as a Theater of Voices: Disorganization and Reorganization of a Position Repertoire. *Journal of Constructivist Psychology*, *19*(2), 147–169. https://doi.org/10.1080/10720530500508779

Hinshelwood, R. D. (1997). The elusive concept of "internal objects" (1934-1943): Its role in the formation of the Klein group. *The International Journal of Psychoanalysis*, *78*(5), 877.

Horvath, A. O. (2011). Alliance in Common Factor Land: A view through the research lens. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *14*(1), 121. https://doi.org/10.4081/ripppo.2011.45

Hurry, S., Matteo, C., Ayodeji, O., & Goswami, E. (2020). Spitch's robo-bankers are learning Swiss German! *IMD Switzerland*.

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

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of Dialogue in Infancy: Coordinated Timing in Development. *Monographs of the*

*Society for Research in Child Development*, *66*(2,), 1–149.

Joshi, A., Kumar, M., & Das, P. K. (2016). Speaker Diarization: A review. *2016 International Conference on Signal Processing and Communication (ICSC)*, 191–196. https://doi.org/10.1109/ICSPCom.2016.7980574

Jung, E., Pick, O., Schlüter-Müller, S., Schmeck, K., & Goth, K. (2013). Identity development in adolescents with mental problems. *Child and Adolescent Psychiatry and Mental Health*, *7*(1), 26. https://doi.org/10.1186/1753-2000-7-26

Kaess, M., Parzer, P., Mattern, M., Plener, P. L., Bifulco, A., Resch, F., & Brunner, R. (2013). Adverse childhood experiences and their impact on frequency, severity, and the individual function of nonsuicidal self-injury in youth. *Psychiatry Research*, *206*(2–3), 265–272. https://doi.org/10.1016/j.psychres.2012.10.012

Kernberg, O. F. (1995). *Object relations theory and clinical psychoanalysis*. Jason Aronson.

Kernberg, O. F. (2006). Identity: Recent Findings and Clinical Implications. *The Psychoanalytic Quarterly*, *75*(4), 969–1004. https://doi.org/10.1002/j.2167-4086.2006.tb00065.x

Kernberg, O. F., Yeomans, F. E., Clarkin, J. F., & Levy, K. N. (2008). Transference focused psychotherapy: Overview and update. *The International Journal of Psychoanalysis*, *89*(3), 601–620. https://doi.org/10.1111/j.1745-8315.2008.00046.x

Kleinbub, J. R. (2017). State of the Art of Interpersonal Physiology in Psychotherapy: A Systematic Review. *Frontiers in Psychology*, *8*, 2053. https://doi.org/10.3389/fpsyg.2017.02053

Koole, S. L., & Tschacher, W. (2016). Synchrony in Psychotherapy: A Review and an Integrative Framework for the Therapeutic Alliance. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00862

Kring, A. M. (2008). Emotion Disturbances as Transdiagnostic Processes in Psychopathology in Handbook of Emotions. In *Handbook of Emotions* (Third Ed., p. 691). New Guilford Press.

Laroche, J., Berardi, A. M., & Brangier, E. (2014). Embodiment of intersubjective time: Relational dynamics as attractors in the temporal coordination of interpersonal behaviors and experiences. *Frontiers in Psychology*, *Vol. 5*, 17. https://doi.org/10.3389/fpsyg.2014.01180

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), 467–481. https://doi.org/10.1037/a0034332

Levy, K. N., Clarkin, J. F., Yeomans, F. E., Scott, L. N., Wasserman, R. H., & Kernberg, O. F. (2006). The mechanisms of change in the treatment of borderline personality disorder with transference focused psychotherapy. *Journal of Clinical Psychology*, *62*(4), 481–501. https://doi.org/10.1002/jclp.20239

Lifshitz, K., & Blair, J. H. (1960). The polygraphic recording of a repeated hypnotic abreaction with comments on abreaction psychotherapy. *Journal of Nervous and Mental Disease*, *130*, 246–252. https://doi.org/10.1097/00005053-196003000-00010

Lind, M., Vanwoerden, S., Penner, F., & Sharp, C. (2019). Inpatient Adolescents With Borderline Personality Disorder Features: Identity Diffusion and Narrative Incoherence. *Personality Disorders: Theory, Research, and Treatment*, 6. http://dx.doi.org/10.1037/per0000338

Linehan, M. M., Cochran, B. N., Mar, C. M., Levensky, E. R., & Comtois, K. A. (2000). Therapeutic burnout among borderline personality disordered clients and their therapists: Development and evaluation of two adaptations of the maslach burnout inventory. *Cognitive and Behavioral Practice*, *7*(3), 329–337. https://doi.org/10.1016/S1077-7229(00)80091-7

Lingiardi, V., & Colli, A. (2015). Therapeutic Alliance and Alliance Ruptures and Resolutions: Theoretical Definitions, Assessment Issues, and Research Findings. In *Psychotherapy Research* (pp. 311–329). Springer. https://doi.org/10.1007/978-3-7091-1382-0_16

Lyons-Ruth, K. (1999). The two-person unconscious: Intersubjective dialogue, enactive relational representation, and the emergence of new forms of relational organization. *Psychoanalytic Inquiry*, *19*(4), 576–617. https://doi.org/10.1080/07351699909534267

Mac Cormack, T. (1997). Believing in Make-Believe: Looking at Theater as a Metaphor for Psychotherapy. *Family Process*, *36*(2), 151–169. https://doi.org/10.1111/j.1545-5300.1997.00151.x

Martinez, C., Tomicic, A., & Medina, L. (2012). Dialogic Discourse Analysis of Psychotherapeutic Dialogue: Microanalysis of relevant Psychotherapy Episodes. *International Journal for Dialogical Science*, *Vol. 6*(No. 1), 99–121.

McMain, S. F., Boritz, T. Z., & Leybman, M. J. (2015). Common strategies for cultivating a

positive therapy relationship in the treatment of borderline personality disorder. *Journal of Psychotherapy Integration*, *25*(1), 20–29. https://doi.org/10.1037/a0038768

Mitchell, S. A., & Aron, L. E. (1999). *Relational psychoanalysis: The emergence of a tradition.* Analytic Press.

Muran, J. C., & Safran, J. D. (2017). Therapeutic alliance ruptures. In *The sage encyclopedia of abnormal and clinical psychology: Vol. Vol. 1*. SAGE Publications.

Muran, J Christopher. (2002). A relational approach to understand change: Plurality and contextualism in a psychotherapy research program. *Psychotherapy Research*, *12*(2), 113–138.

Muran, J. Christopher, & Safran, J. D. (2002). A Relational Approach to Psychotherapy. In *Comprehensive handbook of psychotherapy: Psychodynamic/object relations: Vol. Vol. 1* (pp. 253–281). John Wiley & Sons.

Norcross, J. C., & Hill, C. E. (2002). Empirically supported therapy relationships. *Psychotherapy Relationships That Work: Therapist Contributions and Responsiveness to Patients*.

Norcross, J. C., & Wampold, B. E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology*, *67*(2), 127–132. https://doi.org/10.1002/jclp.20764

Normandin, L., Ensink, K., & Kernberg, O. F. (2015). Transference-Focused Psychotherapy for Borderline Adolescents: A Neurobiologically Informed Psychodynamic Psychotherapy. *Journal of Infant, Child, and Adolescent Psychotherapy*, *14*(1), 98–110. https://doi.org/10.1080/15289168.2015.1006008

Owens, L. F. (1977). *An investigation of eye movements and representational systems*.

Pascual-Leone, A., & Yeryomenko, N. (2017). The client "experiencing" scale as a predictor of treatment outcomes: A meta-analysis on psychotherapy process. *Psychotherapy Research*, *27*(6), 653–665. https://doi.org/10.1080/10503307.2016.1152409

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. https://doi.org/10.1037/int0000099

Peluso, P. R., & Freund, R. R. (2018). Therapist and client emotional expression and psychotherapy outcomes: A meta-analysis. *Psychotherapy*, *55*(4), 461–472.

https://doi.org/10.1037/pst0000165

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03). https://doi.org/10.1017/S0954579405050340

Prince, R. (1997). Is talk cheap or silence golden: The negotiation of activity in psychotherapy. *Psychoanalytic Inquiry*, *17*(4), 549–558. https://doi.org/10.1080/07351699709534148

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), 284–295. https://doi.org/10.1037/a0023419

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. https://doi.org/10.1521/jscp.2014.33.5.481

Rice, L. N., & Greenberg, L. S. (1984). *Patterns of change: Intensive analysis of psychotherapeutic process*. Guilford.

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

Roessler, R., & Lester, J. W. (1976). Voice predicts affect during psychotherapy. *Journal of Nervous and Mental Disease*.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, *6*(3), 412.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and Vocal Expressions of Emotion. *Annual Review of Psychology*, *54*(1), 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Safran, J. D. (1993). Breaches in the therapeutic alliance: An arena for negotiating authentic

relatedness. *Psychotherapy: Theory, Research, Practice, Training*, *30*(1), 11–24. https://doi.org/10.1037/0033-3204.30.1.11

Safran, J. D., & Muran, J. C. (1996). The Resolution of Ruptures in the Therapeutic Alliance. *Journal of Consulting and Clinical Psychology*, *Vol. 64*(No. 3), 447–458.

Safran, J. D., Segal, Z. V., Hill, C., & Whiffen, V. (1990). Refining strategies for research on self-representations in emotional disorders. *Cognitive Therapy and Research*, *14*(2), 143–160. https://doi.org/10.1007/BF01176206

Salvatore, S., Tschacher, W., Gelo, O. C. G., & Koch, S. C. (2015). Editorial: Dynamic systems theory and embodiment in psychotherapy research. A new look at process and outcome. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00914

Schenk, N., Rosset, V., Inderbinen, M., Fürer, L., Schmeck, K., & Schlüter-Müller, S. (2020). Alliance Ruptures in der Psychotherapie von Jugendlichen mit Borderline-Persönlichkeitspathologie: Gefahr oder Chance? *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *69*(1), 60–81. https://doi.org/10.13109/prkk.2020.69.1.60

Schenk, N., Zimmermann, R., Fürer, L., Krause, M., Weise, S., Kaess, M., Schlüter-Müller, S., & Schmeck, K. (2019). Trajectories of alliance ruptures in the psychotherapy of adolescents with borderline personality pathology: Timing, typology and significance. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *22*(2). https://doi.org/10.4081/ripppo.2019.348

Schiepek, G., & Strunk, G. (2010). The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—A method for the real-time monitoring of human change processes. *Biological Cybernetics*, *102*(3), 197–207. https://doi.org/10.1007/s00422-009-0362-1

Schlüter-Müller, S., Birkhölzer, M., Jung, E., & Schmeck, K. (2020). Adolescent Identity Treatment bei Persönlichkeitsstörungen im Jugendalter: Eine integrative Therapiemethode. *Psychotherapeut*, *65*(5), 383–389. https://doi.org/10.1007/s00278-020-00447-5

Schlüter-Müller, S., Goth, K., Jung, E., & Schmeck, K. (2015). Assessment and Treatment of Identity Pathology During Adolescence. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, *3*(1), 63–70. https://doi.org/10.21307/sjcapp-2015-006

Schmeck, K., Fegert, J. M., & Schlüter-Müller, S. (2013). On identity. *Child and Adolescent Psychiatry and Mental Health*, *7*(24).

Sharma, G., & Dhall, A. (2021). A Survey on Automatic Multimodal Emotion Recognition in the Wild. In *Advances in Data Science: Methodologies and Applications*. Springer. https://doi.org/10.1007/978-3-030-51870-7_3

Sharp, C., Wright, A. G. C., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology*, *124*(2), 387–398. https://doi.org/10.1037/abn0000033

Silberschatz, G. (2017). Improving the yield of psychotherapy research. *Psychotherapy Research*, *27*(1), 1–13. https://doi.org/10.1080/10503307.2015.1076202

Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., & Staiger, P. K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, *57*, 141–163. https://doi.org/10.1016/j.cpr.2017.09.002

Sollberger, D. (2013). On identity: From a philosophical point of view. *Child and Adolescent Psychiatry and Mental Health*, *7*(1), 29. https://doi.org/10.1186/1753-2000-7-29

Sollberger, D., Gremaud-Heitz, D., Riemenschneider, A., Agarwalla, P., Benecke, C., Schwald, O., Küchenhoff, J., Walter, M., & Dammann, G. (2015). Change in Identity Diffusion and Psychopathology in a Specialized Inpatient Treatment for Borderline Personality Disorder: Change in Identity Diffusion. *Clinical Psychology & Psychotherapy*, *22*(6), 559–569. https://doi.org/10.1002/cpp.1915

Sollberger, D., Gremaud-Heitz, D., Riemenschneider, A., Küchenhoff, J., Dammann, G., & Walter, M. (2012). Associations between Identity Diffusion, Axis II Disorder, and Psychopathology in Inpatients with Borderline Personality Disorder. *Psychopathology*, *45*(1), 15–21. https://doi.org/10.1159/000325104

Soma, C. S., Baucom, B. R. W., 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

Steppan, M., Zimmermann, R., Fürer, L., Schenk, N., & Schmeck, K. (2020). Machine Learning Facial Emotion Recognition in Psychotherapy Research. A useful approach? *PsyArXiv*.

Timulak, L. (2010). Significant events in psychotherapy: An update of research findings. *Psychology and Psychotherapy: Theory, Research and Practice*, *83*(4), 421–447. https://doi.org/10.1348/147608310X499404

Toomin, M. K., & Toomin, H. (1975). GSR biofeedback in psychotherapy: Some clinical observations. *Psychotherapy: Theory, Research & Practice*, *12*(1), 33–38. https://doi.org/10.1037/h0086402

Tschacher, W., & Storch, M. (2012). Die Bedeutung von Embodiment für Psychologie und Psychotherapie. *Psychotherapie in Psychiatrie, Psychotherapeutischer Medizin und Klinischer Psychologie*, *17*(2), 259–267.

Tufekcioglu, S., & Muran, J. C. (2014). *A relational approach to personality disorder & alliance rupture* (p. 25). Guilford.

Vogt, T., André, E., & Bee, N. (2008). EmoVoice—A Framework for Online Recognition of Emotions from Voice. In E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, & M. Weber (Eds.), *Perception in Multimodal Dialogue Systems* (Vol. 5078, pp. 188–199). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-69369-7_21

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.

Watkins, J. C., & Watkins, H. H. (1997). *Ego states: Theory and therapy*. WW Norton.

Whissell, C. (2009). Using the Revised Dictionary of Affect in Language to Quantify the Emotional Undertones of Samples of Natural Language. *Psychological Reports*, *105*(2), 509–521. https://doi.org/10.2466/PR0.105.2.509-521

Wilkinson-Ryan, T., & Drew, W. (2000). Identity Disturbance in Borderline Personality Disorder: An Empirical Investigation. *American Journal of Psychiatry*, *157*(4), 528–541. https://doi.org/10.1176/appi.ajp.157.4.528

Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C., & Narayanan, S. S. (2015). “Rate My Therapist”: Automated Detection of Empathy in Drug and Alcohol Counseling via Speech and Language

Processing. *PLOS ONE*, *10*(12), e0143055. https://doi.org/10.1371/journal.pone.0143055

Yadav, S. K., Bhushan, M., & Gupta, S. (2015). Multimodal sentiment analysis: Sentiment analysis using audiovisual format. *IEEE*, 5.

Yeomans, F. E., Levy, K. N., & Caligor, E. (2013). Transference-focused psychotherapy. *Psychotherapy*, *50*(3), 449–453. https://doi.org/10.1037/a0033417

Zanarini, M. C., Weingeroff, J. L., & Frankenburg, F. R. (2009). Defense Mechanisms Associated With Borderline Personality Disorder. *Journal of Personality Disorders*, *23*(2), 113–121. https://doi.org/10.1521/pedi.2009.23.2.113

Zhang, A., Wang, Q., Zhu, Z., Paisley, J., & Wang, C. (2019). Fully Supervised Speaker Diarization. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6301–6305. https://doi.org/10.1109/ICASSP.2019.8683892

Zhang, H., Yan, X., & Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, *94*, 1–10. https://doi.org/10.1016/j.autcon.2018.05.033

Zilcha-Mano, S. (2016). New analytic strategies help answer the controversial question of whether alliance is therapeutic in itself. *World Psychiatry*, *15*(1), 84–85. https://doi.org/10.1002/wps.20305

Zilcha-Mano, S. (2017). Is the Alliance Really Therapeutic? Revisiting This Question in Light of Recent Methodological Advances. *American Psychologist*, *72*(4), 311–325. http://dx.doi.org/10.1037/a0040435

Zilcha-Mano, S. (2019). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research*, *29*(6), 693–708. https://doi.org/10.1080/10503307.2018.1429691

Zilcha-Mano, S., & Errázuriz, P. (2017). Early development of mechanisms of change as a predictor of subsequent change and treatment outcome: The case of working alliance. *Journal of Consulting and Clinical Psychology*, *85*(5), 508–520. https://doi.org/10.1037/ccp0000192

Zilcha-Mano, S., McCarthy, K. S., Dinger, U., & Barber, J. P. (2014). To What Extent Is Alliance Affected by Transference? An Empirical Exploration. *Psychotherapy*, 1–10.

http://dx.doi.org/10.1037/a0036566